AL/HR-TR-1994-0107

AD-A285 450

ARMSTRONG LABORATORY

AN EMPIRICAL APPROACH TO VISUAL DISPLAY PREFERENCE BASED UPON MODULATION TRANSFER FUNCTION AND LUMINANCE

Ronald J. Evans

University of Dayton Research Institute
300 College Park Avenue
Dayton, OH 45469-0110

HUMAN RESOURCES DIRECTORATE
AIRCREW TRAINING RESEARCH DIVISION
6001 South Power Road, Building 558
Mesa, AZ 85206-0904

August 1994

Final Technical Report for Period September 1992 - July 1993

Approved for public release; distribution is unlimited.

AIR FORCE MATERIEL COMMAND
BROOKS AIR FORCE BASE, TEXAS

NOTICES

This technical report is published as received and has not been edited by the technical editing staff of the Armstrong Laboratory.

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

The Office of Public Affairs has reviewed this report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This report has been reviewed and is approved for publication.

Elizabeth L. Martin
ELIZABETH L. MARTIN
Project Scientist

Dee H. Andrews
DEE H. ANDREWS, Technical Director
Aircrew Training Research Division

Lynn A. Carroll
LYNN A. CARROLL, Colonel, USAF
Chief, Aircrew Training Research Division

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	August 1994	Final – September 1992 – July 1993

4. TITLE AND SUBTITLE
An Empirical Approach to Visual Display Preference Based Upon Modulation Transfer Function and Luminance

5. FUNDING NUMBERS
C – F33615-90-C-0005
PE – 62205F
PR – 1123
TA – 03
WU – 85

6. AUTHOR(S)
Ronald J. Evans

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Dayton Research Institute
300 College Park Avenue
Dayton, OH 45469-0110

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAMES(S) AND ADDRESS(ES)
Armstrong Laboratory (AFMC)
Human Resources Directorate
Aircrew Training Research Division
6001 S. Power Road, Bldg 558
Mesa, AZ 85206-0904

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
AL/HR-TR-1994-0107

11. SUPPLEMENTARY NOTES

Armstrong Laboratory Technical Monitor: Dr. Elizabeth Martin, (602) 988-6561.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The purpose of the present study was to develop a three-dimensional preference space for displays as a function of the display modulation transfer function (MTF) and average display luminance. For any MTF-luminance combination, then, the goal was to generate a point in the third dimension denoting the preference for that pair. A paired comparison experiment was conducted where, on individual trials, observers viewed side-by-side images varying in MTF (5 levels) and average luminance (4 levels). The 5 X 4 = 20 combinations of MTF and luminance could be thought of as 20 filters. Preferences on individual trials were cumulated into empirical preference probability matrices which denoted the probability of preferring any one of the 20 filters over any of the 20 filters. A psychological model of preference, the Bradley-Terry-Luce (BTL) Model was then fit to the matrices in order to estimate a scale value or preference for each of the 20 filters or points in the three-dimensional space. Regression techniques (R^2 =.98) were used to generate a preference surface in the three-dimensional space, from which the preference for any display could be predicted. Additional analyses indicated that not only did ratings differ significantly based upon changes in MTF and luminance, but ratings change significantly ($P < .001$) as a function of using different scenery in the images. Finally, in predictive equations generated from the data, changes in MTF area (measured in percent contrast x cycles per degree of visual angle) tended to have about three times as much effect on preference as did changes in luminance (measured in footlamberts).

14. SUBJECT TERMS
Bradley-Terry-Luce Model, Contrast, Gaussian MTF, Image quality, Modulation transfer function, MTF, Multidimensional scaling, Paired comparisons, Resolution

15. NUMBER OF PAGES
86

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

CONTENTS

List of Figures

List of Figures (concluded)

List of Tables

PREFACE

The present effort was conducted in support of the Armstrong Laboratory/Aircrew Training Research Division (AL/HRA) research concerning image quality in simulator displays. The goal of this effort was to develop a quantitative model which predicted display quality based upon the display modulation transfer function and display luminance. The project was conducted under Work Unit 1123-03-85, Flying Training Research Support. Research support was provided by the University of Dayton Research Institute under Contract No. F33615-90-C-0005. The contract monitor was Ms. Patricia Spears, AL/HRAP.

The goal of this specific research effort was to (a) initiate an experimental design and methodology whereby image quality of visual displays could be studied in a multidimensional environment and (b) collect data and develop a multidimensional model to predict image quality as a function of multiple display attributes.

The author wishes to express thanks to Ms. Marge Keslin for final edit of the manuscript.

AN EMPIRICAL APPROACH TO VISUAL DISPLAY PREFERENCE BASED UPON MODULATION TRANSFER FUNCTION AND LUMINANCE

INTRODUCTION

The importance and prevalence of electronic visual displays in society continue to grow as technologies (e.g., fiber optics, microchip and compact disk technology) improve. As computational power and the capability to store and transmit dynamic visual imagery improve, the use of electronic displays will increase. Given the variety of tasks for which humans interact with visual displays and the economics involved in these emerging technologies, the need to develop methods for assessing the quality of imagery and acceptable standards becomes a priority.

The Assessment of Display Quality as a Complex Task

Electronic visual displays vary along a number of dimensions, including luminance, resolution, contrast, display size, color, and update rate. When comparing the quality of two displays, it may be necessary to compare multiple dimensions or attributes listed above. In many instances, improvement along one dimension is often accompanied by a loss in another dimension. For example, increases in display size or field of view are typically accompanied by a drop in luminance. Systems with greater luminance capabilities may exhibit a drop in their resolution at high spatial frequencies. Such trade-offs occur rather frequently with electronic visual displays. These trade-offs will often make choices unclear. For example, larger displays viewed from farther away may have more resolution but less brightness than smaller displays.

Figure 1 shows the Michelson Contrast as a function of spatial frequency, often simply referred to as the display modulation transfer function (MTF), for two displays used at the Aircrew Training Research Division of Armstrong Laboratory in Mesa, AZ. Functions based upon the height and area under the MTF curve often serve as image quality metrics (e.g., Barten, 1989, 1991; Evans, 1990). Note how the curves of the two displays cross over at approximately 2-3 cycles per degree on the x-axis. Using only these two curves for comparison, it might be difficult to determine which display is preferred. Although the limited field-of-view display (LFOVD) is physically larger, the display for advanced research and training (DART) is much brighter and is typically preferred.

Referring to Figure 1 as an example, the display MTF is a measure of Michelson contrast (i.e., (Lmax-Lmin)/(Lmax+Lmin) where Lmax and Lmin denote the maximum and minimum luminance measured on the device) capability as a function of spatial frequency. As such, the MTF is a multidimensional display factor in its own right. In the development of image quality metrics, however, the

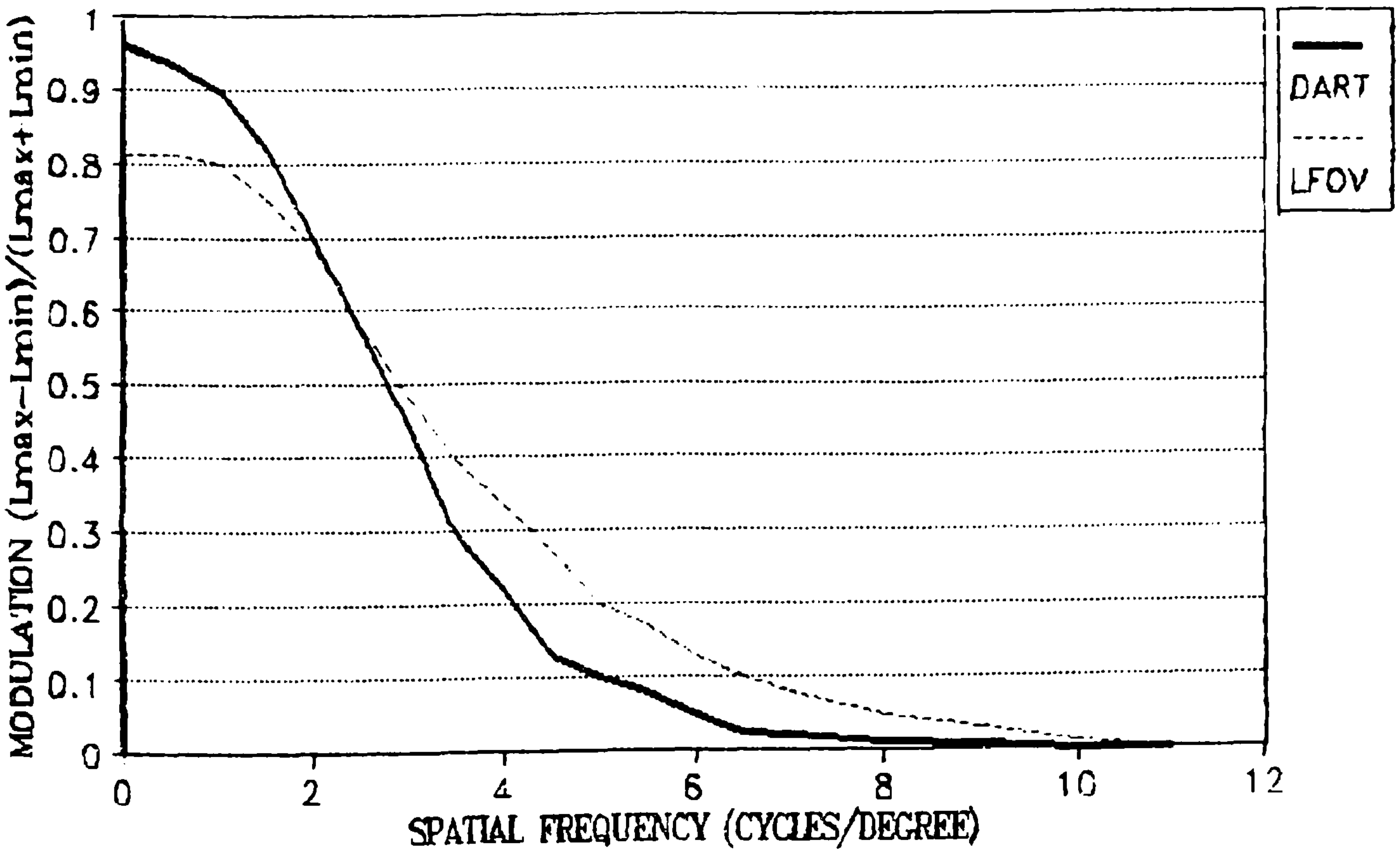


Figure 1
Comparison of Michelson Contrast for DART and LFOV Displays

display MTF has been the most often used display factor to date. For example, in Figure 1 a simple metric would be to compute the area under each curve as a measure of display quality. More advanced metrics (e.g., Barten, 1990; Snyder, 1985) weight the distance along the x-axis differentially. Using Barten's square root integral (SQRI) metric, for example, the DART Display in Figure 1 would be judged to be superior because the integration of the MTF is done with respect to the log of the spatial frequency, thereby emphasizing the lower spatial frequencies. By simply observing the two display MTFs in Figure 1, however, there appears to be no clear-cut winner. The display factor of luminance plays a much greater role in this comparison.

Because of trade-offs similar to those discussed above, a multidimensional approach to assessing image quality is required. Here multiple display factors are manipulated and observer preference or performance is measured. Such an approach allows us to relate changes across dimensions of interest (i.e., luminance, display size) by their effect on the dependent variable in the experiment.

There are problems with using such an approach, however. From a hardware viewpoint, it is quite difficult to manipulate display factors or parameters independently of one another. If image quality is compared across two display systems, the difficulty is in manipulating the factors of interest while holding other display factors equivalent across the two displays. The alternative used in the present study is to use a single display and manipulate display parameters within the single display. As will be seen in the present study, use of a single display limits the range of the variables manipulated in the study.

In the present study, two factors have been chosen to be manipulated. These are the display MTF (actually, the area under the MTF curve) and the display luminance. The goal of the present study is to develop a prediction scheme for display preference as a function of display MTF and luminance. Figure 2 portrays a plane or two-dimensional graph where display MTF and luminance represent the x and y axes. If all displays can be characterized as points in this plane, then a third orthogonal dimension can be used to represent preference. Note that in Figure 2 both MTF and luminance are unidimensional quantities. For the present study, the area under the display MTF will be used to represent the unidimensional MTF quantity and average display luminance will be used to represent luminance. Both of these definitions are simplifications of the actual concepts concerning contrast, resolution, and brightness within imagery.

Up to this point, the discussion of display quality has focused only upon the characteristics of the display. Figure 3

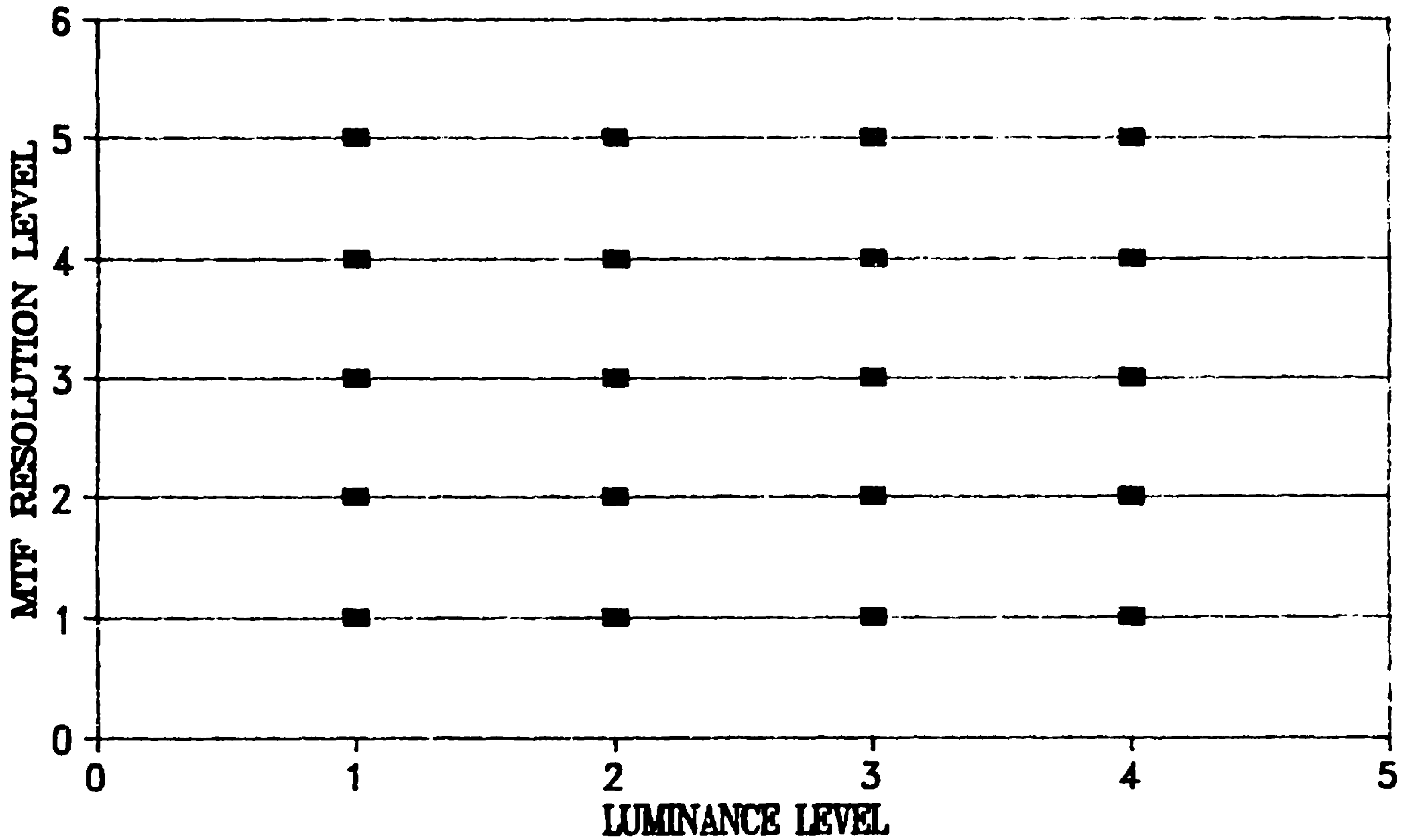


Figure 2
Two-Dimensional Plot of MTF and Luminance Combinations

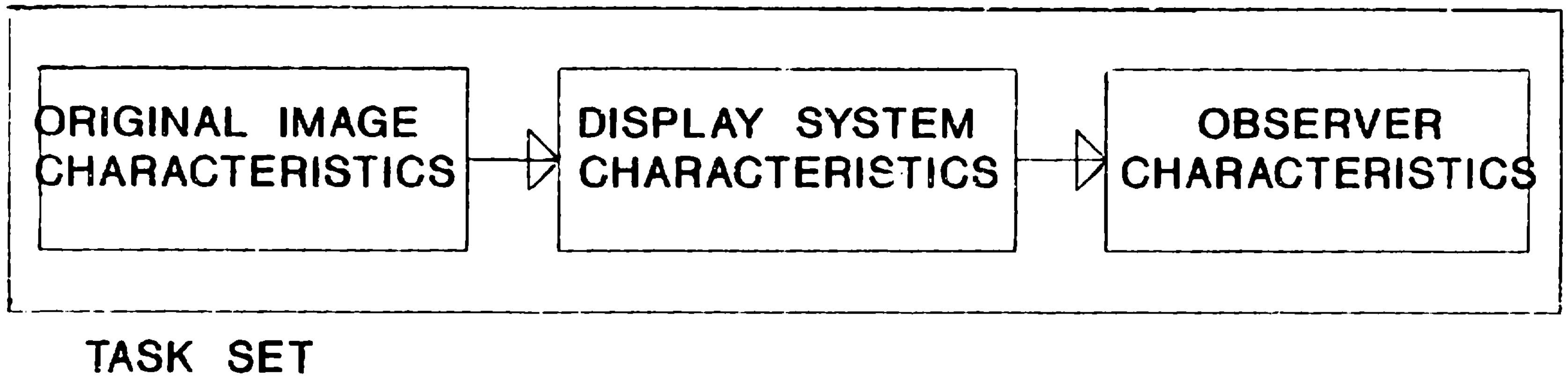


Figure 3
Systems Approach to Image and Display Quality

portrays a systems approach to image quality which emphasizes (a) the characteristics of the image which are fed to the display, (b) the effect of the display system on the imagery, and (c) the perceptual processing and limitations of the observer. In Figure 3, these three components are contained within a box defined as the task domain. Theoretically, the quality of a display could be determined for asymptotic levels of (a), (c), and the task domain. For the purpose of procuring displays for specific functional tasks, though, the quality of a display cannot be ascertained independently of components (a), (c), or the task domain. For example, consider two displays which are equivalent except that one display's limiting resolution is 25 cycles/degree while the other's limiting resolution is 30 cycles/degree. If the imagery fed to the display is limited to 20 cycles/degree, the two displays will appear to be equivalent for all tasks. As in any practical linear system, the limitation of the weakest component of the system becomes a limiting factor in the system's performance. The entire system is included within the task domain box as a reminder that the imagery or information at the front end of the system and the processing at the observer end will typically be task specific.

An important concept in the systems approach from Figure 3 is that the information contained in the original image affects all subsequent processes. Because of this and the inherent variability in and between images, several images will be used in the present study.

Along with the systems approach, which emphasizes the sequential nature and influence of components on one another, the notion of image quality must be operationalized in order to be studied in a scientific manner. Operational definitions of image quality typically rely on human response or performance which hypothetically varies as a function of the display manipulations. In a majority of previous experimental studies, two task types have been primarily used in defining image quality. In the first type of task, observers rate imagery from displays based upon aesthetic qualities (e.g., sharpness with Kusaka, 1989). This method of defining image quality may involve global interpretation (i.e., based upon a large composite of display factors) by the observer and can be quite subjective. The second type of task uses performance measures such as target detection, recognition, and identification (e.g., Snyder, 1985). In these tasks, improvements in performance would be interpreted as an improvement in image quality. Manipulations in rating tasks are usually easily detectable and the experimenter is not interested in working near the bounds of the observer's perceptual limits. Performance tasks typically employ threshold manipulations of factors in order to produce errors or have some deleterious effect on performance.

In the current study, a paired-comparisons rating task has been chosen as the dependent measure. On individual trials, observers will be presented with two images simultaneously, both images being filtered by two of a number of MTF-luminance combinations. The images will be identical except for the application of the two filters. Observer responses will denote which of the two images is preferred. Referring back to Figure 2, the purpose of obtaining the paired-comparison preferences will be to generate points in a third dimension denoting preference for each MTF-luminance combination. Modeling will be used to generate these preferences from the paired-comparison preferences and regression will be used to generate a preference surface in the three-dimensional space.

Twenty points (5 MTF levels x 4 luminance levels) will be generated in the plane of Figure 2. Using the paired-comparisons procedure, there will be estimates for the preference of any of the 20 points (MTF-luminance combinations) over any of the 20 points (400 possible comparisons). Included in these comparisons is the pairing of a point with itself. This condition serves as an estimate of noise or observer sensitivity.

As mentioned previously, the MTF and luminance filtering of the imagery is a technically difficult task. In this study, both images are presented side-by-side on a single display. The display itself serves as a limitation in the filtering process. The following section describes how MTF and luminance filtering were accomplished for the study.

TRANSFORMATION OF IMAGE MTF AND LUMINANCE

In order to create the 20 levels of imagery (5 MTFs X 4 luminance levels) required in the experimental design, it was necessary to digitally filter each of the five images by each of the 20 filters. Before discussing how the filtering was accomplished, a short introduction to the original nature of the five digital images is in order. It is emphasized here again that, as shown in Figure 3, the characteristics of the original image play an integral role in any experimental results obtained. The effect of any of the three components in Figure 3 on the quality of the final perception cannot be interpreted in isolation of the remaining two components.

Photographs of the images were digitally scanned into 512 rows X 512 columns of Digital-to-Analog Code (DAC) values. Each of the 512 X 512 DAC values could take any value between 0 and 255 (an eight-bit DAC value). Figure 4 shows a frequency distribution of DAC values for each of the five images (airport scene, crop scene, mountain scene, ocean scene, and pines scene). The y-axis in this graph is a logarithmic transformation of frequency. Note that all the images contain an abundance of low luminance (low DAC values)

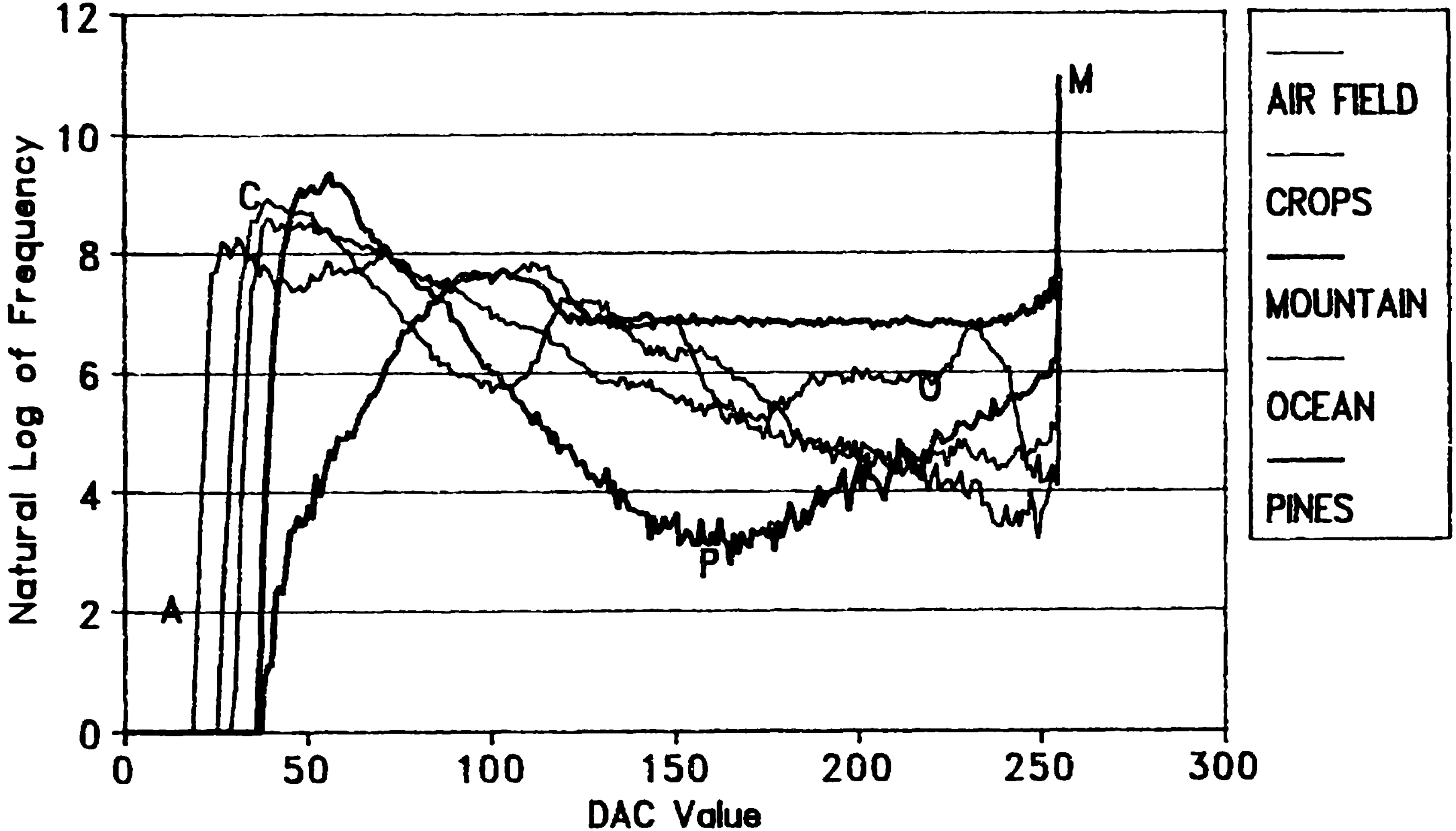


Figure 4
Distribution of DAC Values for Five Images

pixels as well as an abundance of pixels with high luminance (DAC value of 255). The importance of the frequency distributions in this graph will be explained in the section covering luminance filtering of the images. Figure 5 shows a plot of the gamma curve for the display used in the experiment. This curve relates the DAC value of a pixel in the image to the luminance (in fL) from the display used in the experiment (as measured with a photometer). Two curves are plotted in Figure 5. The upper or higher curve shows luminance measured from the display as a function of DAC value measured from a small area in the center of the display when the surrounding area was completely black (BACKGROUND DAC = 0). The lower curve shows similar measurements when the surrounding area was made as bright as possible (BACKGROUND DAC = 255). Note that very little variation in luminance occurs below a DAC value of 50, essentially making this range of DAC values (0-50) unusable as a source of luminance variation. A power curve was generated which fit between the two curves in Figure 5 for DAC values in the range of 50 to 255. The equation for this curve is as follows:

$$\text{Luminance (fL)} = .00043 \times (\text{DAC VALUE})^2. \tag{1}$$

Equation (1) and the curves in Figure 5 are completely dependent upon the brightness setting of the display. The brightness control of the display was held constant throughout the experiment, and the only purpose in the use of the curves and Equation (1) was to relate the digital representation of the images to the luminance produced by the display CRT.

Beyond specifying the luminance profile of the images used in the experiment, another measure often used to describe imagery is the spatial frequency content. Figure 6 shows a one-dimensional (across rows) global Fourier analysis of the digitized images. As will be the case with most natural imagery (e.g., Evans, 1993), the shape of the curves from Figure 6 are highly similar and provide little discriminable information. A Fourier analysis of local areas in these images can yield curves which are highly dissimilar but, when averaging over large spatial areas of natural imagery, the low frequency components will dominate the Fourier analysis in Figure 6 (denoting large spatial areas of little luminance variation).

The five images used in the experiment are shown in Figures 7 through 11. The images represent an airport scene (Figure 7), a crop scene (Figure 8), a desert mountain scene (Figure 9), an ocean scene (Figure 10), and a hilly pine scene (Figure 11). Note the distinctiveness of these scenes as contrasted with the lack of distinction in the global fourier analysis of the imagery (Figure 6). The following two sections describe the MTF filtering and the luminance filtering applied to the images used in the experiment.

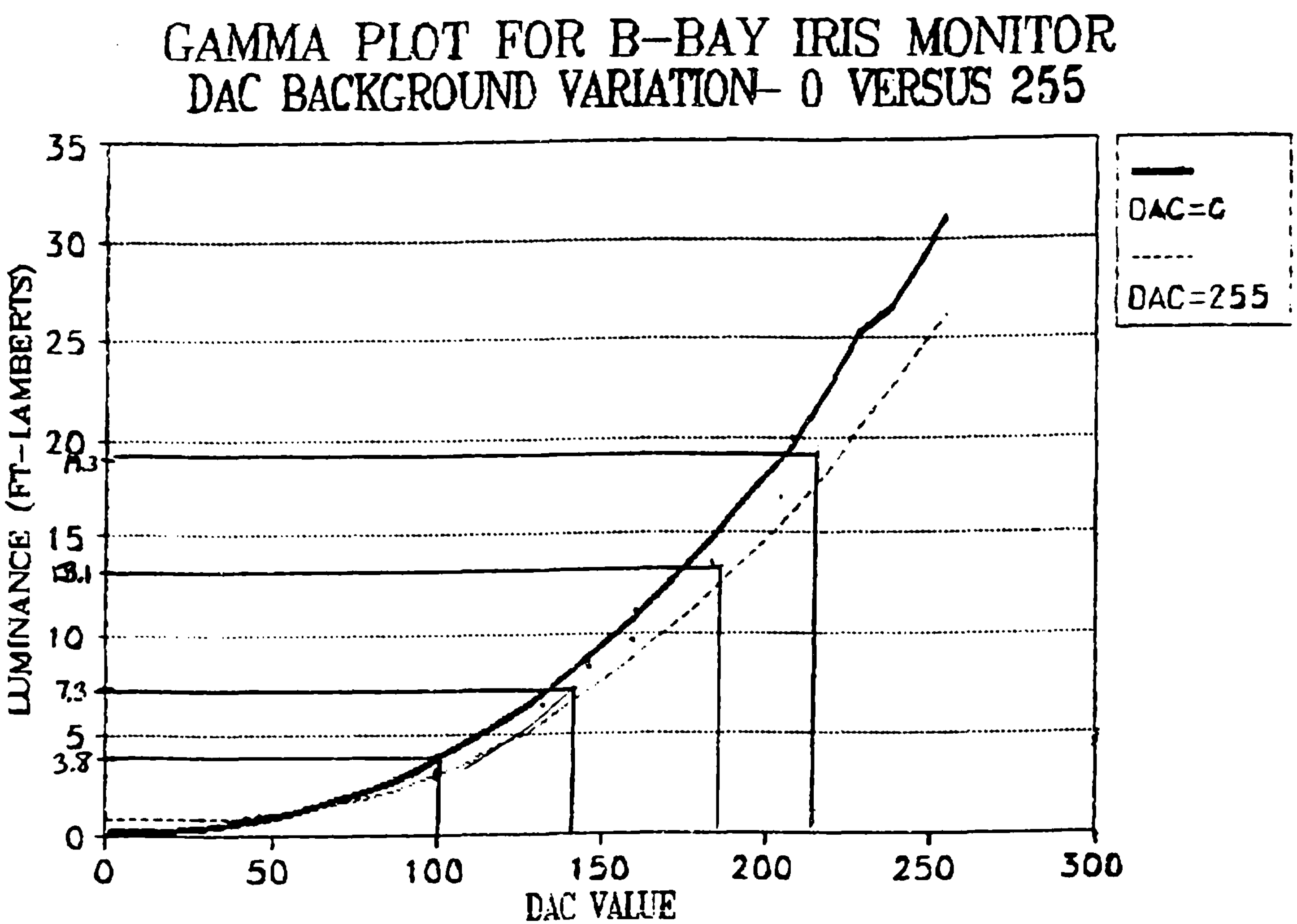


Figure 5
Gamma Curve for the Experimental Display

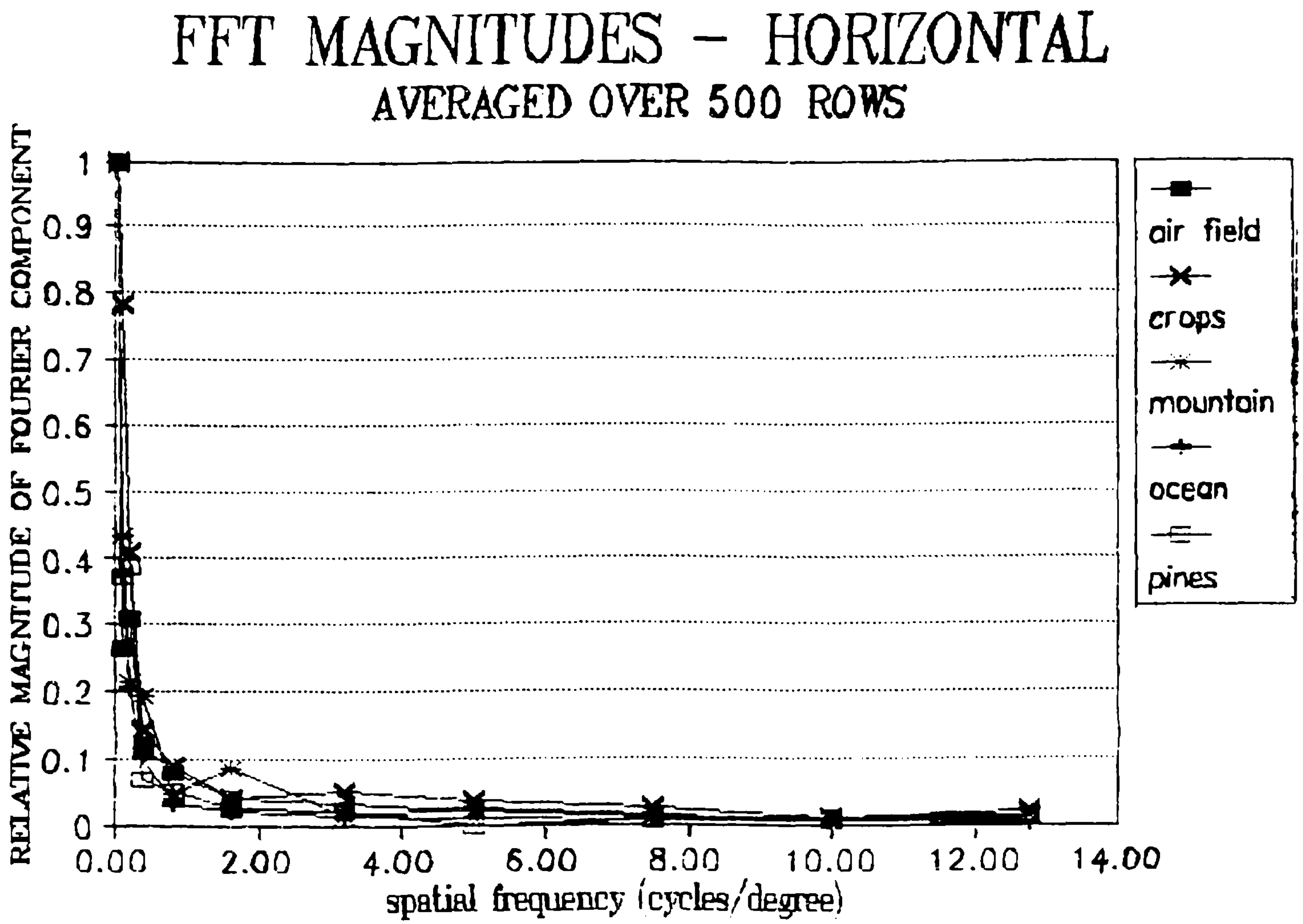


Figure 6
One-Dimensional Fourier Decomposition of Experimental Imagery

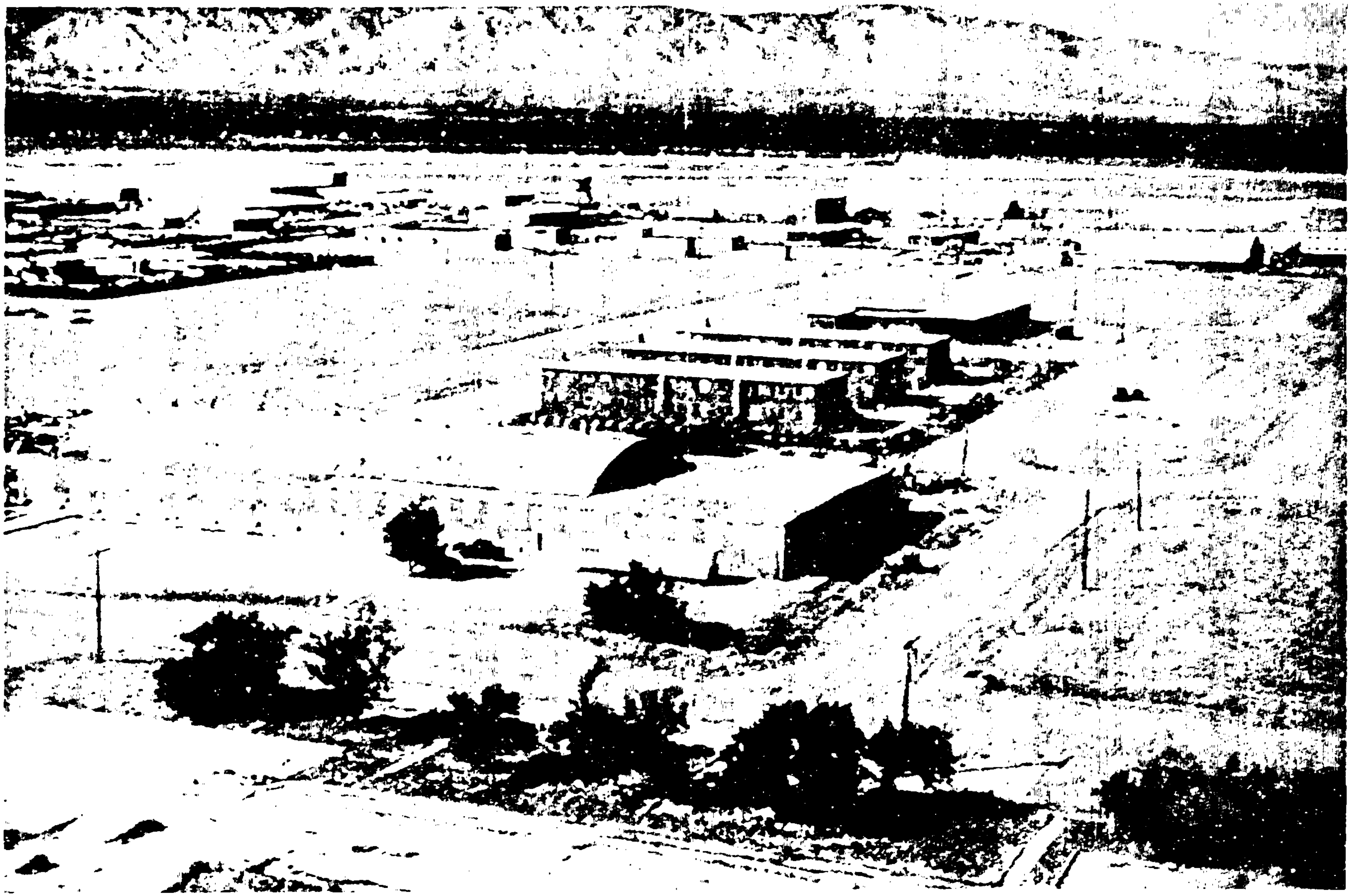
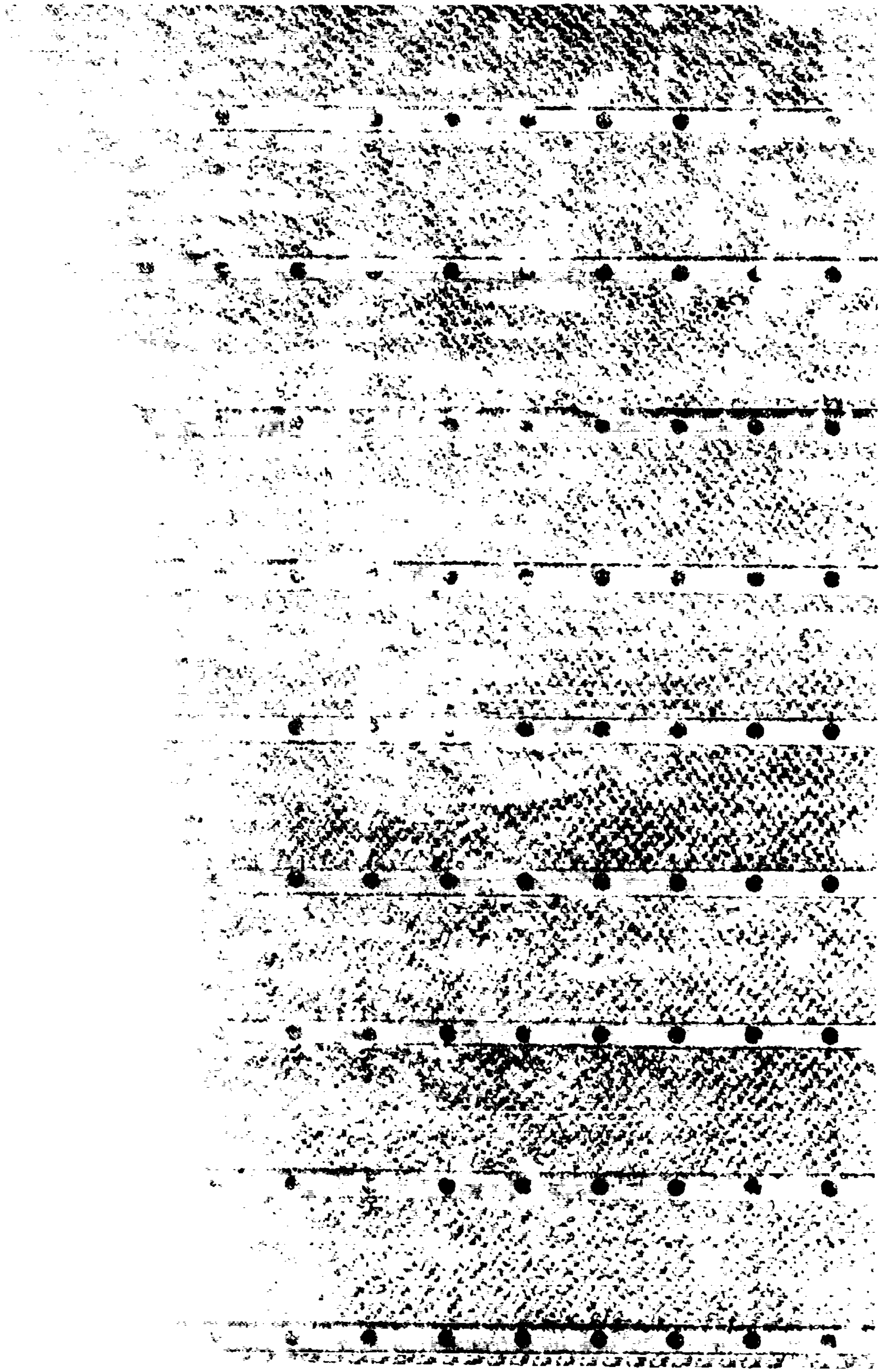


Figure 7
Airport Image

Figure 9
Mou[illegible]



Digital MTF Filtering of Images

The purpose of the MTF filtering of the images is to simulate the effect of a range of display MTFs. In order to do this, the simulated MTFs must (a) look similar to actual display MTFs and (b) be mathematically tractable in order to create and employ the filters on a computer for digital processing. A Gaussian filter is especially attractive for this purpose because it has been linked to CRT display MTFs through physical measures (e.g., Barten, 1984) and its representation in the spatial and frequency domains are relatively simple. In the frequency domain, the equation for the Gaussian MTF is given as:

$$MTF(f) = \exp(-bf^2) \quad \text{(in one dimension)} \quad (2A)$$
$$MTF(f_1, f_2) = \exp(-b(f_1^2+f_2^2)) \quad \text{(in two dimensions)} \quad (2B)$$

f, f_1, and f_2 denote spatial frequency, typically in cycles per degree of visual angle. The rate parameter, b, determines the fall-off of the curve across frequency. Note that there are no multiplicative coefficients preceding the exponential so that at a frequency of zero, the MTF is identically one. The effect of this restriction is that these filters do not change the amount of energy in the signal passing through the system, which is equivalent to specifying that these filters do not change the average luminance of the images.

This representation of a display MTF is problematic. Referring back to Figure 1, neither display's MTF reaches a value of unity at zero spatial frequency. When measuring the Michelson Contrast from a display, the Michelson Contrast will not be unity at zero spatial frequency unless the luminance measurement from the dark screen is identically zero. Usually, there is some amount of ambient light in the environment and even if the display generates no luminance, the ambient illumination will be reflected off of the display. The traditional method of circumventing this shortcoming (i.e., the display MTF < 1 at zero spatial frequency) is to normalize the curve. As can be seen from Figure 1, however, if the two MTFs are normalized, relative information about them is lost. This ambiguity will not be dealt with in the present study but should be kept in mind when comparing actual displays.

In order to filter the 512 x 512 pixel images, the spatial transform of the filters in Equations (2A) and (2B) were used to create convolution filters or kernels (11 x 11) in the spatial domain. These filters were numerically convolved with the images. The spatial filters corresponding to Equations (2A) and (2B) are:

$$h(x) = (\pi/b)^{1/2} \exp(-\pi^2 x^2/b) \quad \text{(in one dimension)} \quad (3A)$$
$$h(x,y) = (\pi/b)^{1} \exp(-\pi^2(x^2+y^2)/b) \quad \text{(in two dimensions)} \quad (3B)$$

where the units of x and y are the inverse of f in Equations 2A and 2B, typically degrees of visual angle. The five MTF filters used in the experiment are shown in Figure 12. The top filter in Figure 12 corresponds to an estimate of the MTF of the IRIS display (1000 vertical x 1024 horizontal pixels). It has been approximated by a Gaussian filter (MTF = $\exp(-.0052X^2)$). Although the Michelson Contrast or MTF at zero spatial frequency was actually .925, the display MTF in Figure 12 has been normalized to a value of one. For a more detailed explanation of the methods used in estimating the MTF of the IRIS display, refer to Evans (1993). The five remaining MTF curves in Figure 12 correspond to a multiplication of the top curve by five curves obtained using Equation (2A) with b set to .015, .03, .05, .074, and .138, respectively.

The reason for multiplying the experimental display MTF by the mathematical MTFs from Equation (2A) is a result of the "double-MTF problem." The image seen by the observer in the experiment is not only filtered by the mathematically generated MTF but also by the display used in the experiment. Using a linear systems approximation, the MTF of the overall system is obtained by taking the product of the MTFs of the two components in the system (i.e., the simulated display and the display used in presenting the images). For a more detailed explanation of the application of the filters to the images, refer to Evans (1993).

Luminance Filtering of Images

The next step in the development of the filtered images was to create four levels of average luminance for the images. Although manipulation of average luminance level is a crude method of studying a factor which is crucial to perception of the image, there currently are no better methods in widespread use.

Each of the 512 x 512 = 262,144 pixels in each image is represented by a DAC value between 0 and 255. Figure 4 shows the distribution of the 262,144 pixels for each of the five images over the 255 possible DAC levels. Figure 5, the gamma curve for the IRIS monitor employed in the experiment, relates the DAC value of the pixels to the luminance viewed on the screen of the monitor. A software program was used to transform the gamma curve into a linear function such that the doubling of a DAC value led to a doubling of luminance. In this manner, it was necessary only to manipulate DAC values in the images in order to perform linear transformations of average luminance.

By examining the images (Figs. 7 through 11) and comparing them with their distributions in Figure 4, a few problems surface with respect to luminance, contrast, and image content. For example, note the pine image with the bright sky. Average luminance in the pine image is dominated on the bright side by the sky (DAC values of 255). The sky provides relatively little

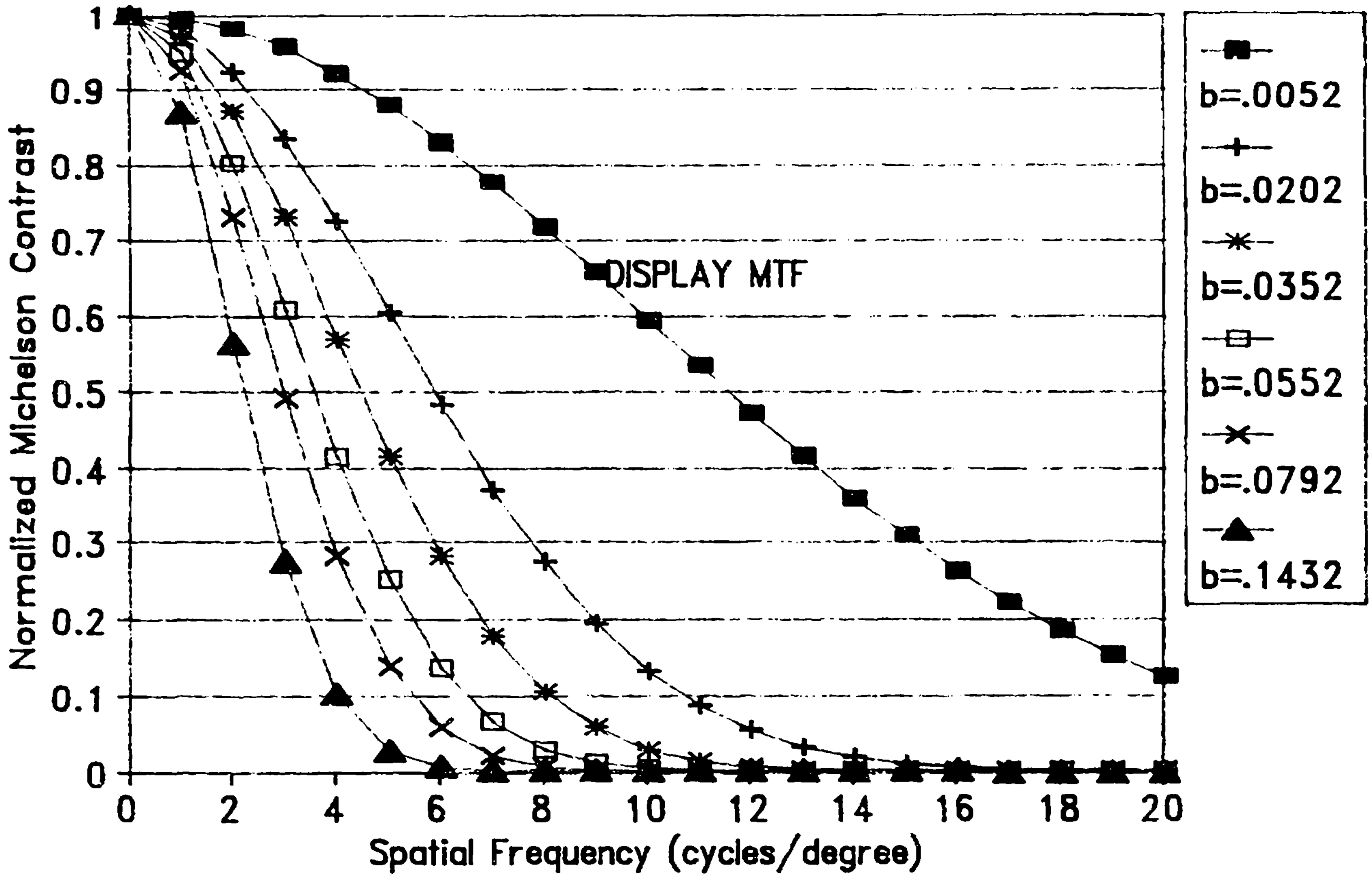


Figure 12
Experimental MTF Filters

information which observers might use in interpreting the image. Therefore, even though the image might be quite bright, most of the information contained in the image (e.g., the outline of the trees), is contained in relatively dark areas of the image. This type of problem highlights the difficulty in describing images (or information content) with respect to variation of specific display parameters (e.g., luminance).

Returning to Figure 4, average luminance in the experiment was manipulated by sliding the frequency distribution to the left or right across the abscissa. Note, however, that if the individual distributions are moved to the left or right, the extreme DAC values 0 and 255 present barriers at both ends of the distribution. The compression process was obtained through a series of trial-and-error manipulations. First, all of the DAC values from an image were divided by a constant to compress the range of the DAC values. Next, a constant was added to each DAC value of an image. The additive constant for each image was set individually so that the average luminance of the images was equal. Figure 13 shows a compression of the DAC distributions of the five images in order that the distributions can be moved across the abscissa to manipulate average luminance. Figures 14 through 18 allow comparison of the original images with the compressed images.

The image compression process highlights the ambiguity of working with image/display/observer components. The compression of the images, as shown in Figures 14 through 18, could be a factor associated with the origin of the image or a factor associated with the display. For experimental purposes, we make the assumption that the compressed images represent the original image content and are not some effect of the display. Experimental results, then, are directly related only to images which have been compressed in the manner shown here.

The compressed images are visually distinct from the original images. In addition, note in Figure 13 that the distribution of DAC values for the Mountain Scene is noticeably shifted to the left relative to the other images even though the average luminance of the images is approximately equal. The spike at the high end of the Mountain Scene represents the sky in that scene and the presence of this spike requires that the entire distribution be shifted to the left to meet the requirement of average equal luminance across images. Although the requirement of average equal luminance across images is one logical constraint, other constraints (e.g., DAC distributions approximately in the same range) may be just as logical. These questions highlight the problem of our inability to characterize the information content of images within some well-specified domain and should be kept in mind in our experimental explorations.

The compressed distributions shown in Figure 13 were moved across the abscissa to create four levels of average luminance,

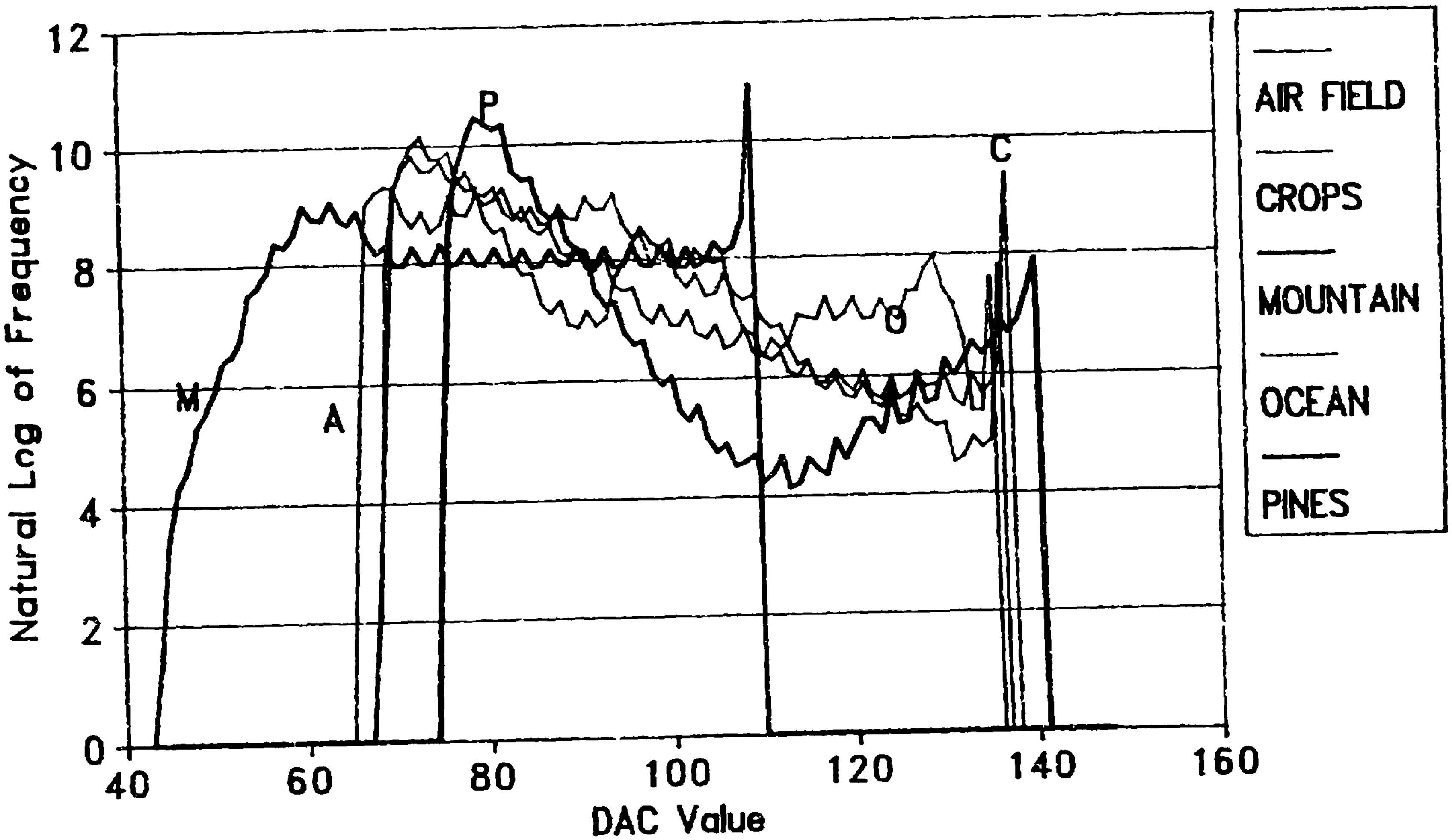
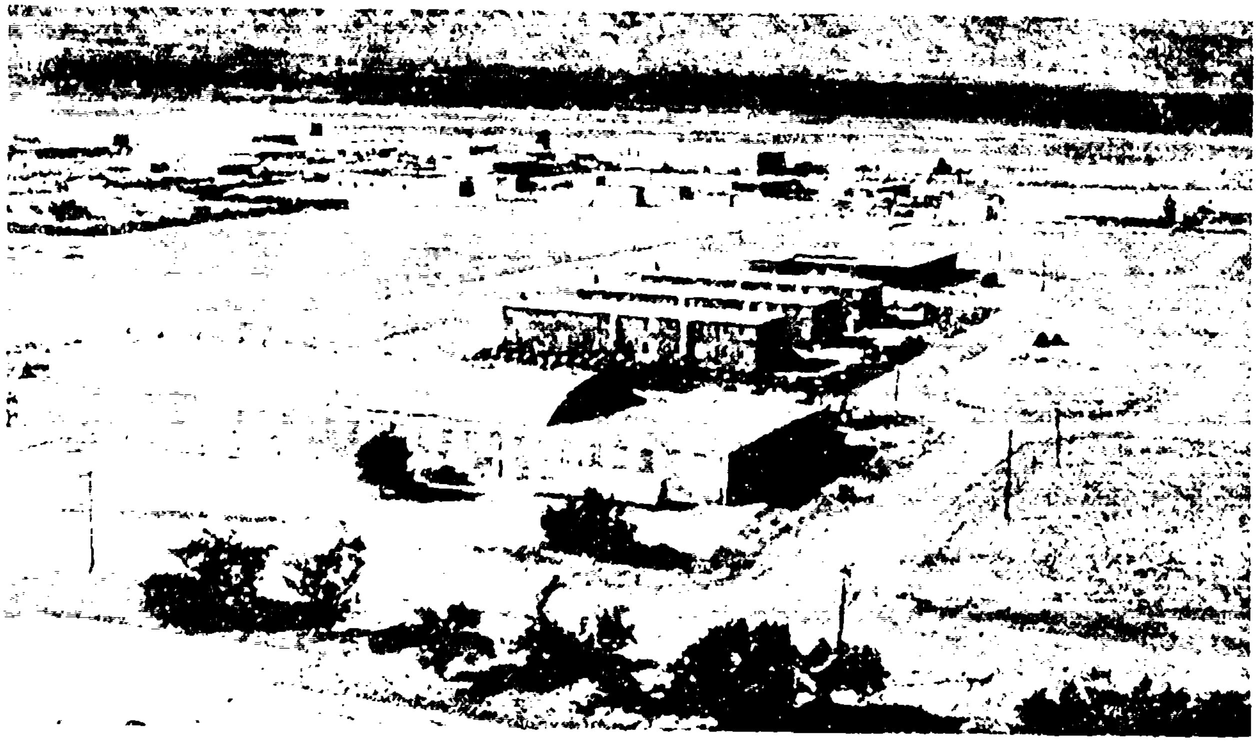


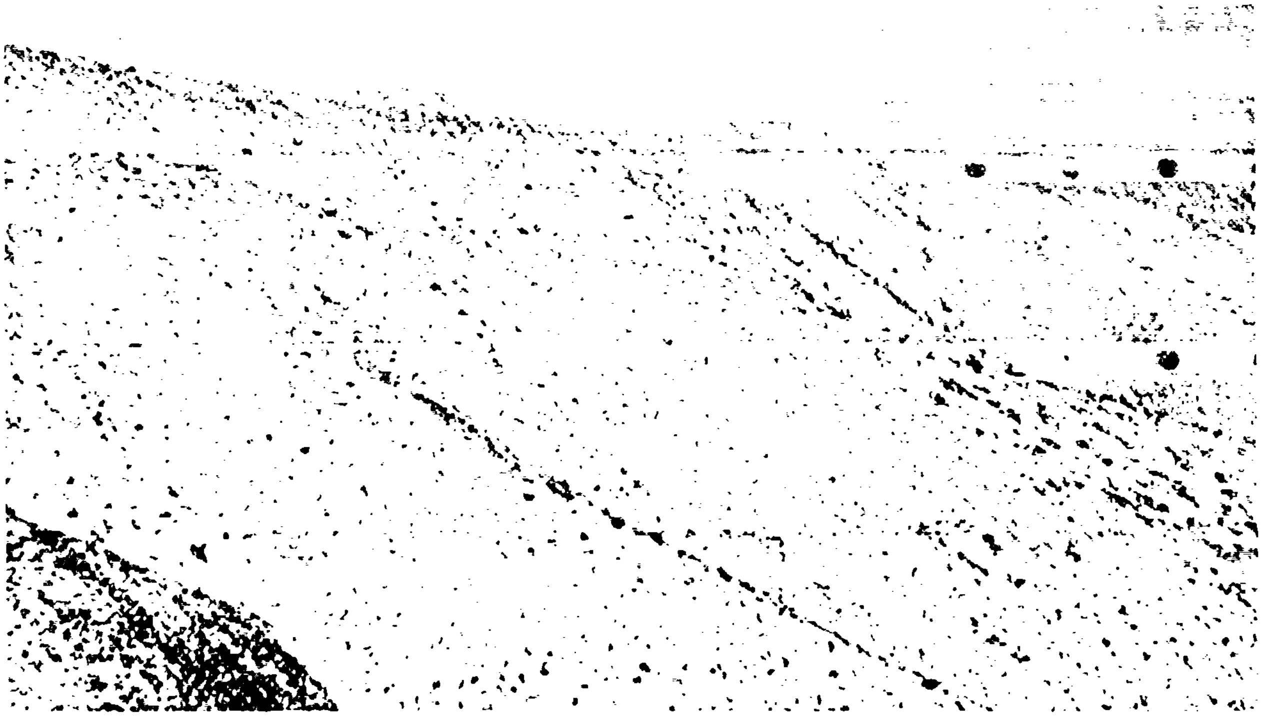
Figure 13
Compressed DAC Distributions for Five Images



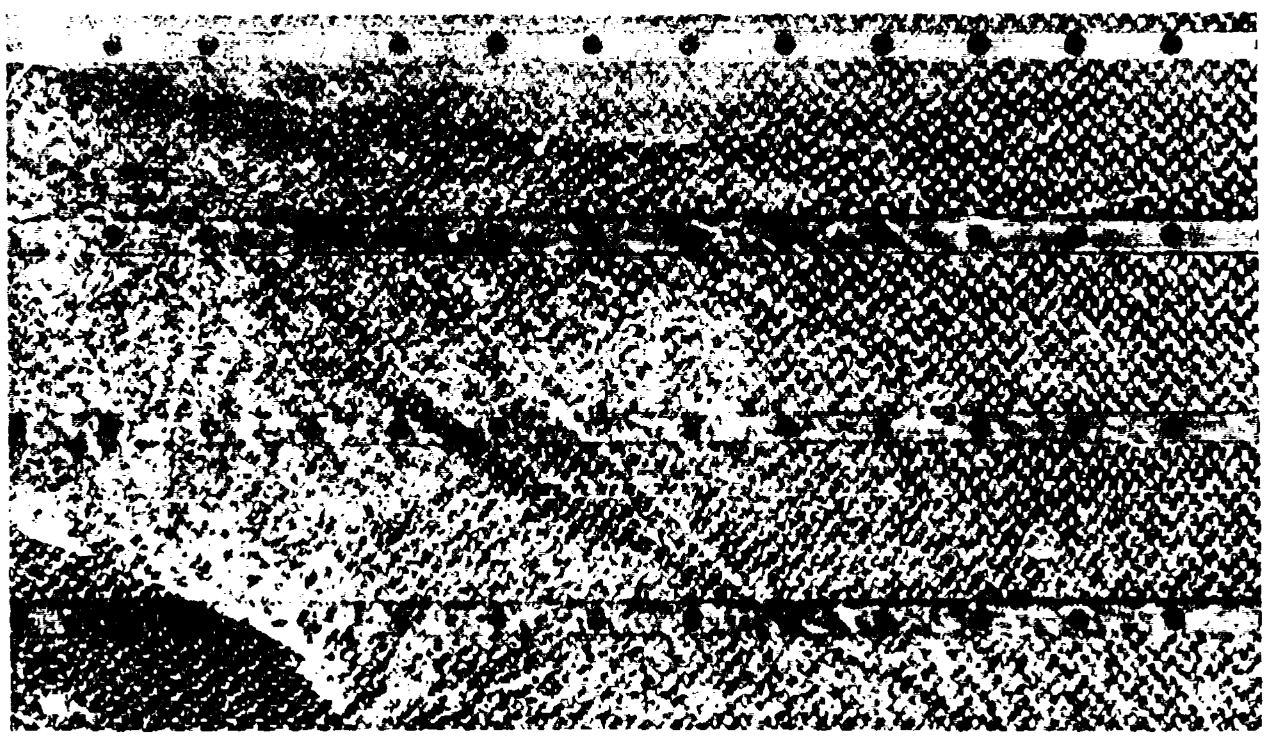
Original

Compressed

Figure 14
Original Versus Compressed Airport Image

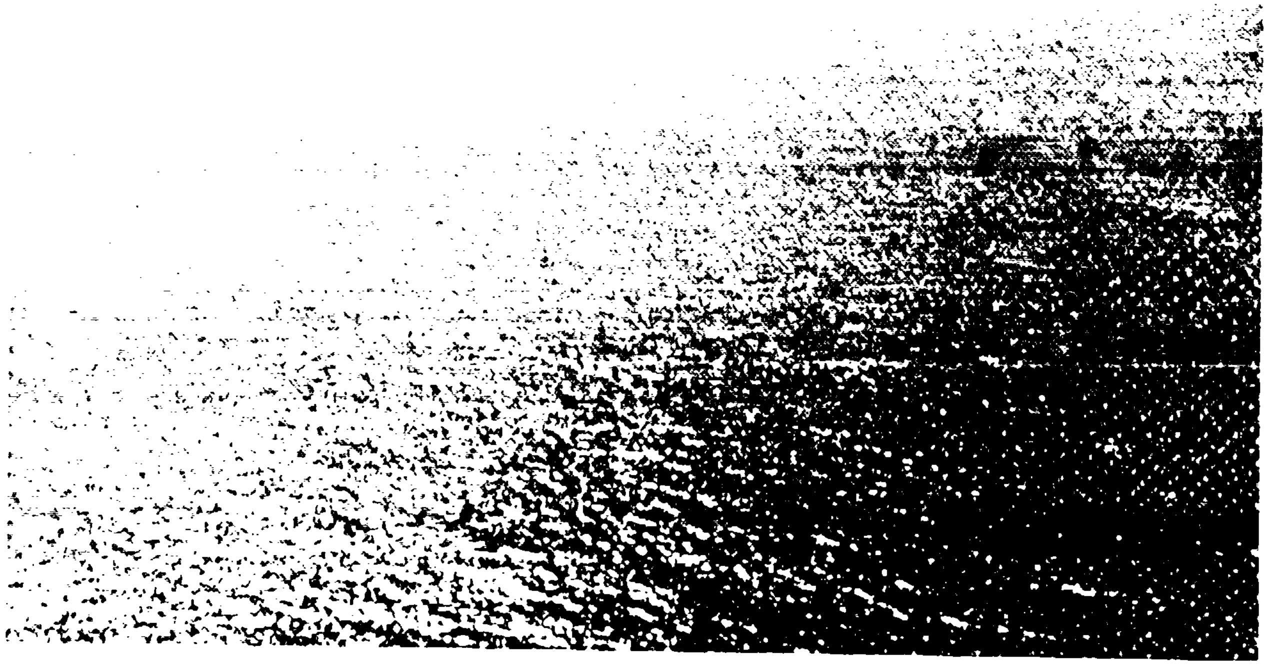


Original



Compressed

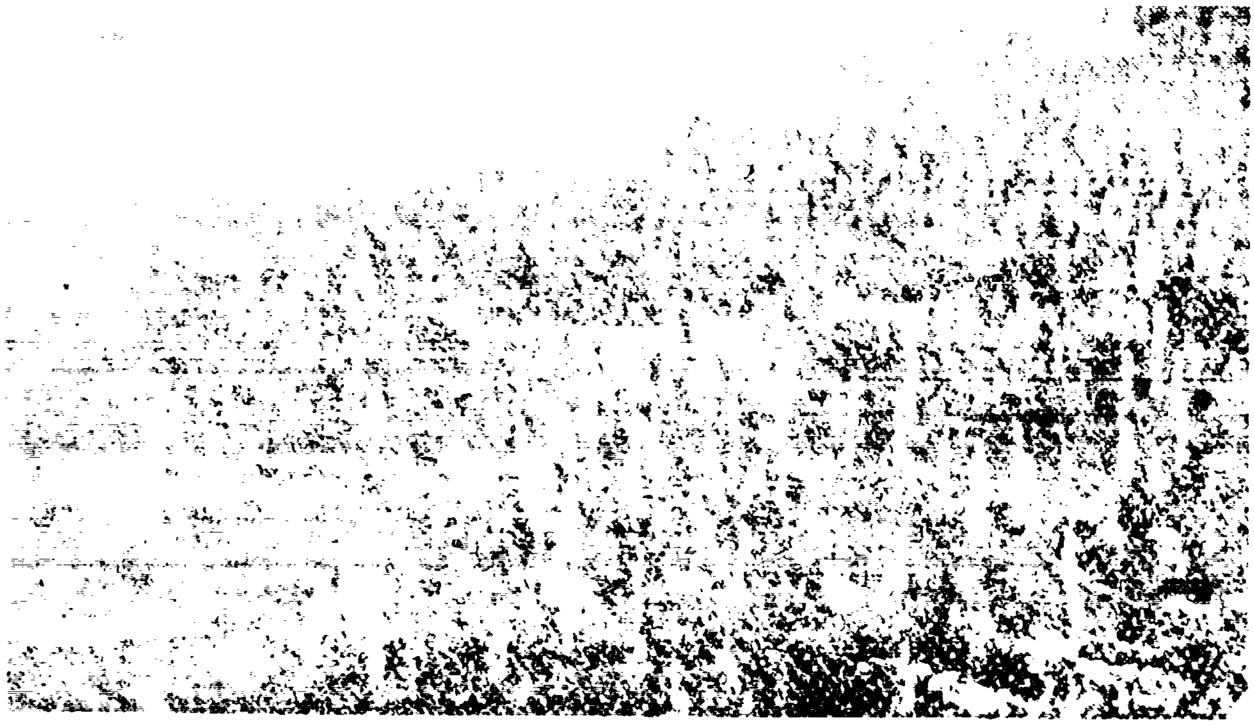
Figure 16
Original Versus Compressed Mountain Image



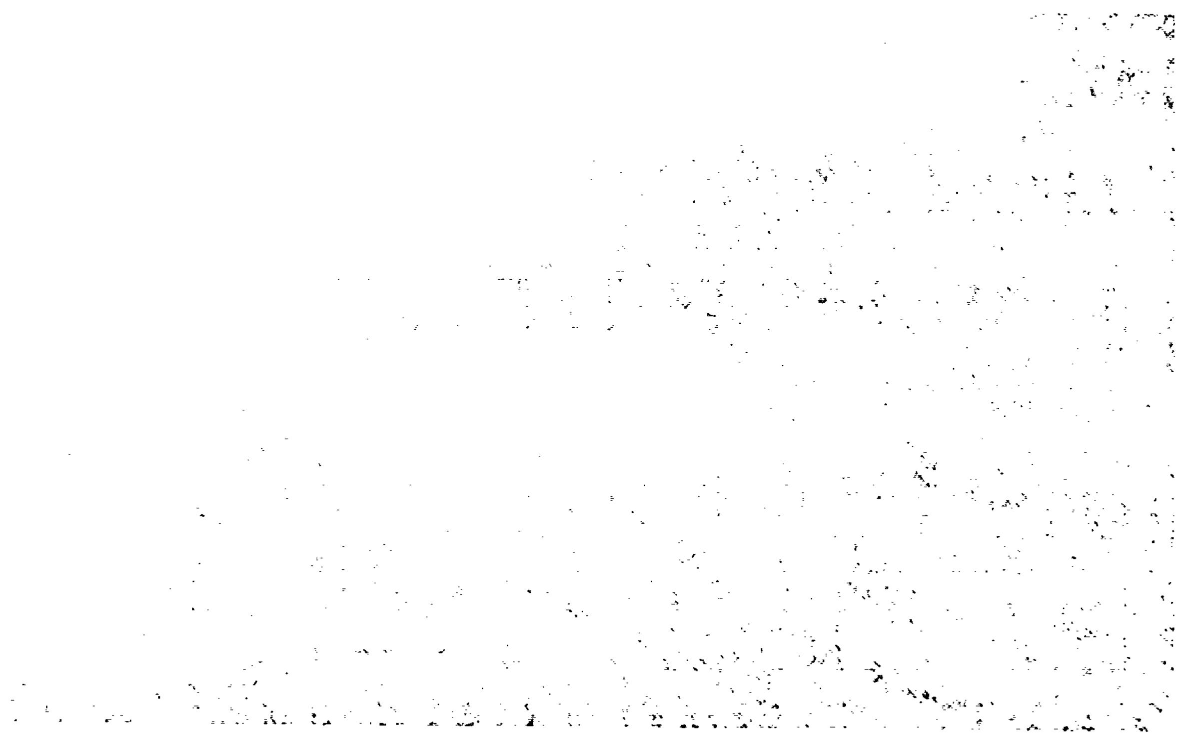
Original

Compressed

Figure 17
Original Versus Compressed Ocean Image



Original



Compressed

Figure 18
Original Versus Compressed Pines Image

approximately 4.0, 6.3, 8.5, and 11.1 fL. Although it was desirable to produce stimuli with average luminances greater than 11.1 fL, in order to do so it would have been necessary to compress the original images more. This artifact demonstrates the limitations in the range of levels of the independent manipulation due to the experimental display device, not only for luminance manipulations but for MTF manipulations as well.

Figures 19 through 23 show three of the five experimental images as filtered by the best, middle, and worst MTFs and two luminance levels only. Due to the photographic and printing processes, these figures are only approximations to the actual experimental stimuli used but should give the reader some perspective of the effect of the 20 filters on the images. In the following section, the methodology of the study is presented in more detail.

EXPERIMENTAL METHODOLOGY

Observers

Six observers, two male and four female, were employed as paid subjects in the experiment. All observers were in their early to mid-twenties.

Apparatus

Figure 24 portrays the experimental setup used. An IRIS graphics workstation with a 15-inch vertical by 15-inch horizontal display was used to display the two images. Each observer placed his/her chin in a chin rest as shown in Figure 24 in order to stabilize viewing. The monitor contained 1,000 vertical by 1,024 horizontal pixels and was viewed from approximately 30 inches by observers. All spatial frequency computations for filtering were performed based upon this 30-inch viewing distance. The entire display subtended approximately 27.6 degrees of visual angle and each pixel subtended approximately 27.6/1024 = .027 degrees or 1.62 minutes of arc. If we assume that the mask between pixels denotes a dark space so that the combination of pixel and mask denotes a light-dark combination, the theoretical limiting resolution of the display would be 1/(.027) or approximately 37 cycles/degree of visual angle. The top curve in Figure 12 shows empirically measured Michelson Contrast for the IRIS monitor as a function of spatial frequency out to only 20 cycles per degree of visual angle.

Observers used a mouse to designate responses to the paired-comparisons presentation. Background illumination from the room was dark, less than .1 fL.

Luminance Filtering

MTF Filter | Bright | Dark

Best

Middle

Worst

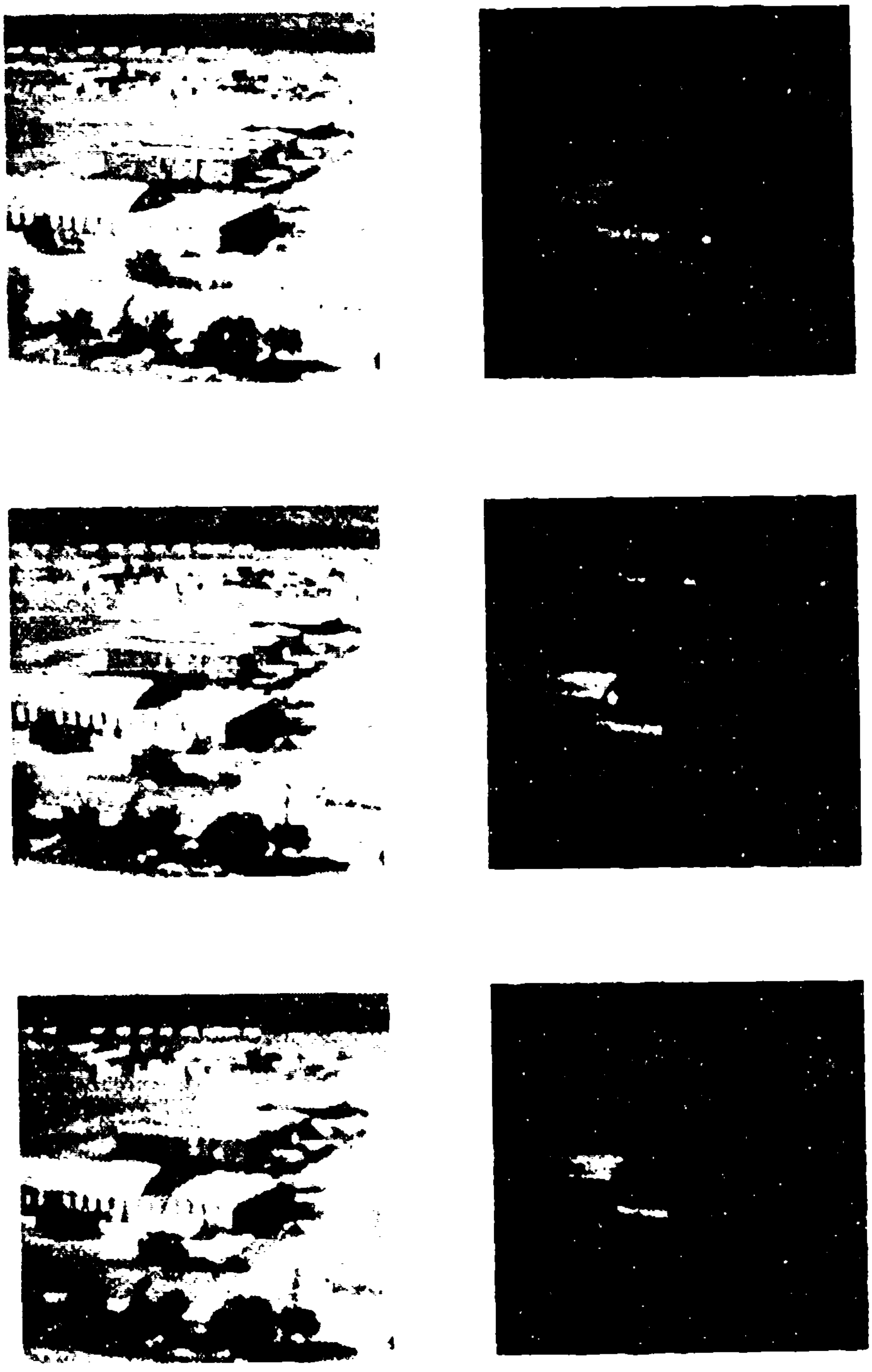


Figure 19
MTF-Luminance Filtering of the Airport Image

Luminance Filtering

MTF Filter | Bright | Dark

Best

Middle

Worst

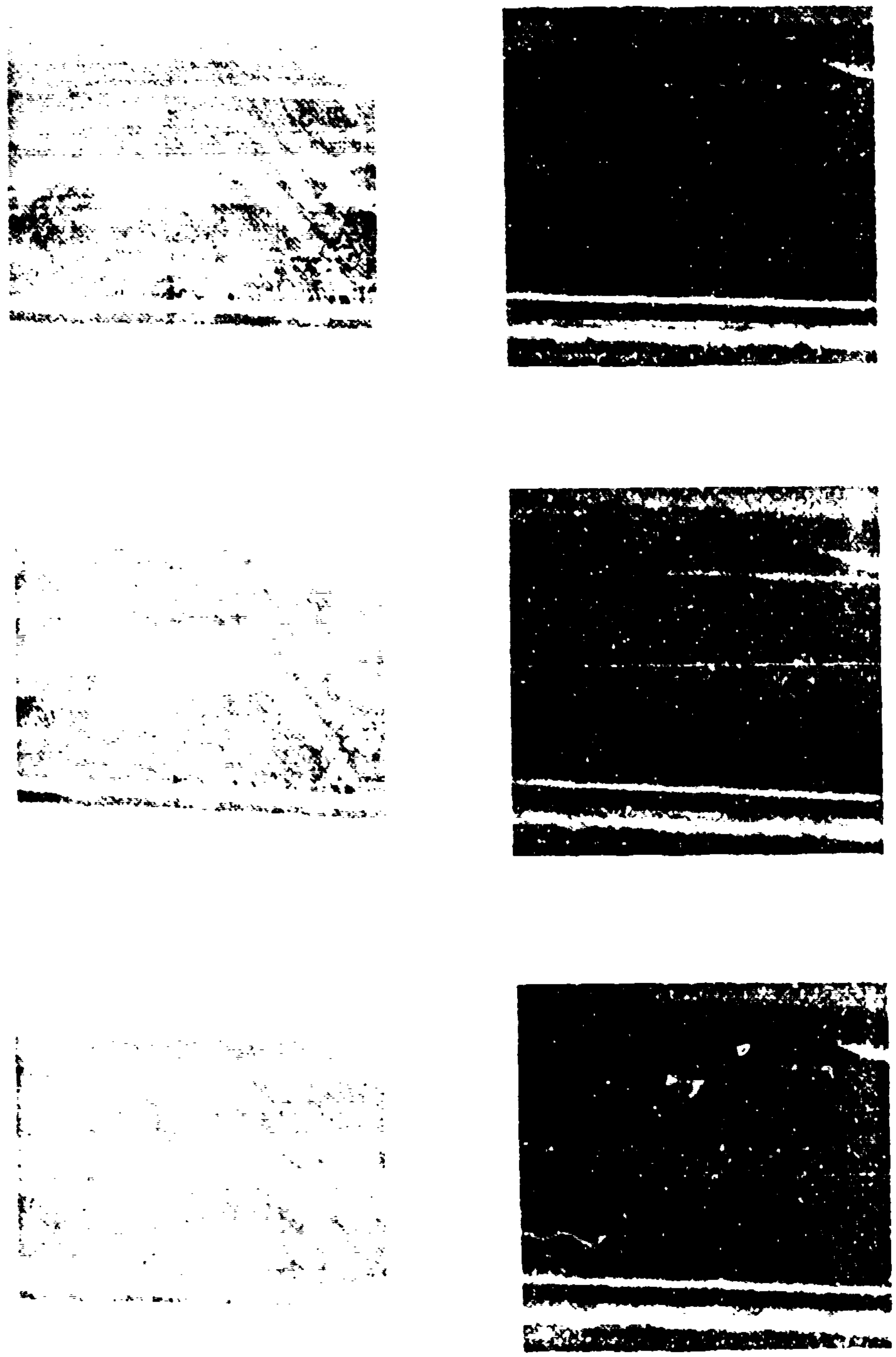


Figure 20
MTF-Luminance Filtering of the Crop Image

Luminance Filtering

MTF Filter	Bright	Dark
Best		
Middle		
Worst		

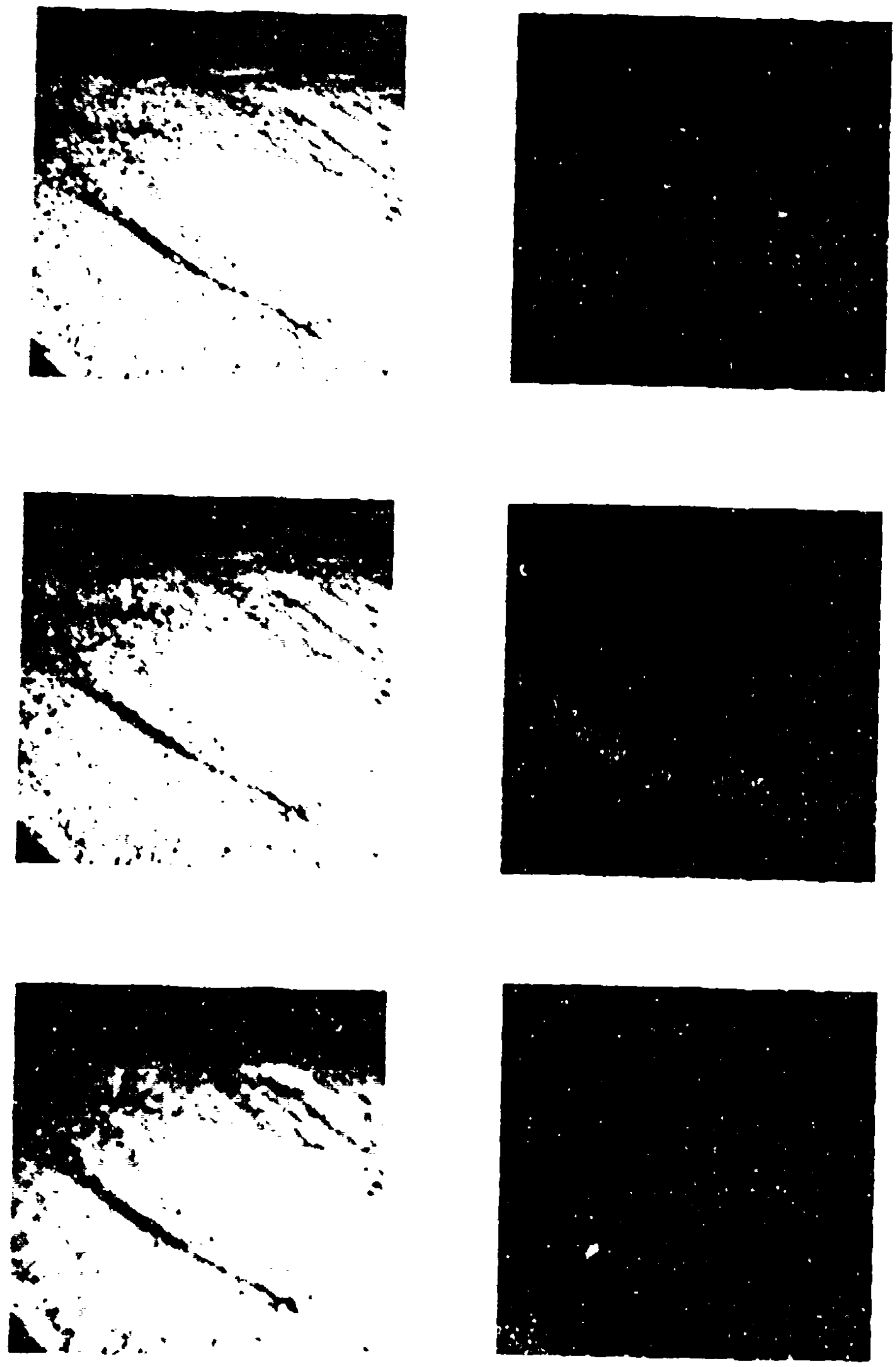


Figure 21
MTF-Luminance Filtering of the Mountain Image

Luminance Filtering

MTF Filter | Bright | Dark

Best

Middle

Worst

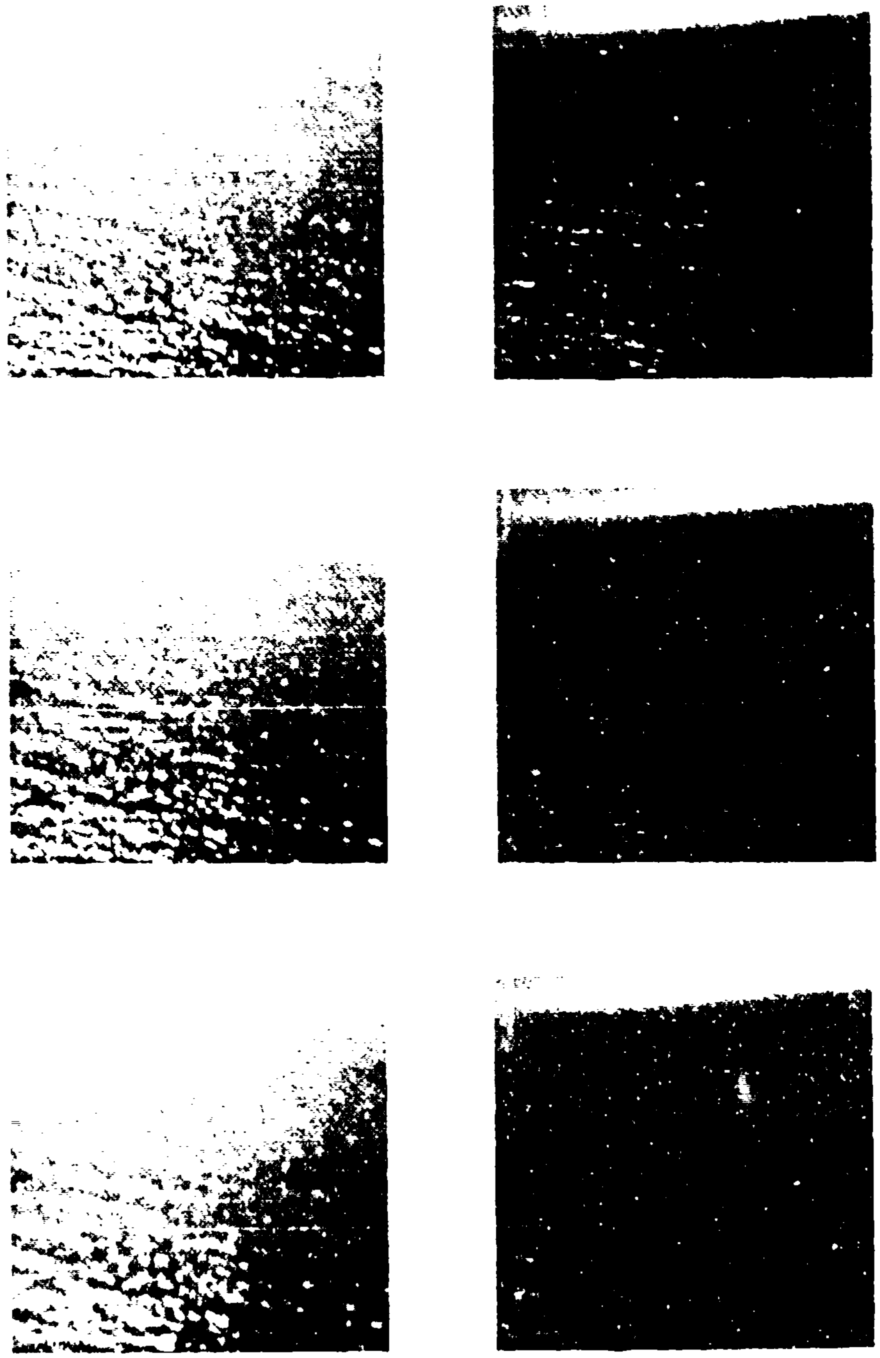


Figure 22
MTF-Luminance Filtering of the Ocean Image

Luminance Filtering

MTF Filter | Bright | Dark

Best

Middle

Worst

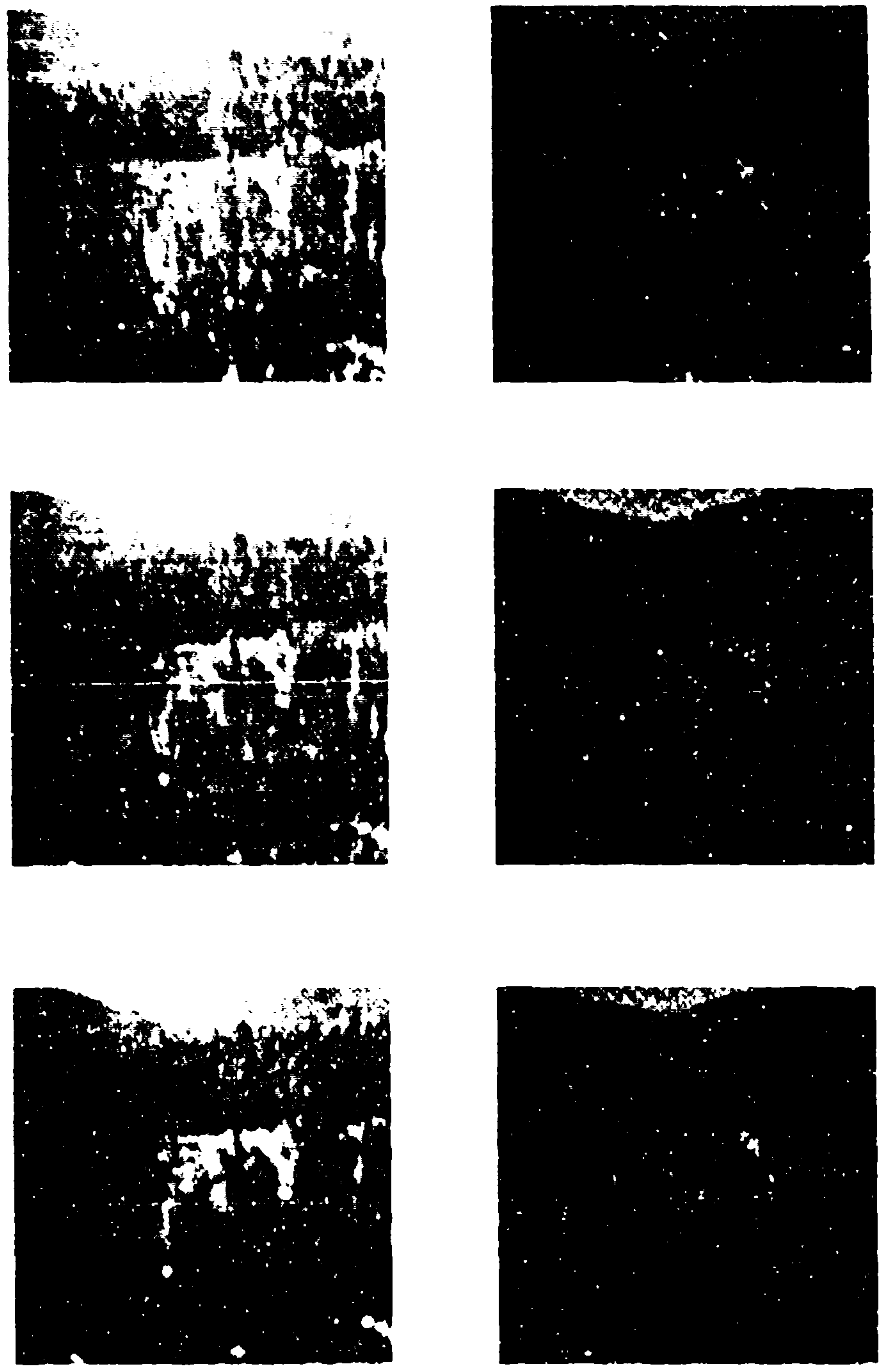


Figure 23
MTF-Luminance Filtering of the Pines Image

Procedure

A trial began with a fixation point in the middle of the screen. Approximately two seconds later, two 512 x 512 images were shown side-by-side in the middle of the display. A small dark patch running vertically in the middle of the screen separated the two images. Since there were only 1,024 pixels spaced horizontally across the screen, only 500 pixels were presented horizontally for each of the two images, leaving a 24-pixel vertical patch separating the two images. The background or luminance surrounding the imagery as it was displayed was approximately 7 fL (DAC = 128). The pre- and post-trial luminances of the screen were less than .1 fL. Both the background and pre/post-trial luminances were considered important constants in the experiment as the contrast of the images against the background both in space and time played a significant role in the visibility of detail in the imagery.

The images remained visible on the screen for approximately two seconds, following which Figure 25 was displayed on the screen. Figure 25 contains a rating scale denoting preference for the left or the right image in the display. The observer used the mouse to slide an arrow across the scale shown on the screen. When the arrow reached the desired location on the rating scale, the observers pressed a button on the mouse to denote their response. A scale value of zero denoted no preference but the observer could also use a second button on the mouse to denote that he/she thought the left and right images were identical. After pressing the button on the mouse, the rating scale disappeared and the observer initiated the next trial by pressing a button on the mouse.

Observers were run through 20 practice and 400 experimental trials each session for 20 sessions, yielding 8,000 experimental trials per observer for the entire study. Each of the five images was presented on 8,000/5 = 1,600 trials so that the 20 x 20 = 400 possible MTF-luminance combinations on each trial were seen 1,600/400 = 4 times for each image.

RESULTS AND DISCUSSION

Data Compilation: Preference Matrices

As shown in Figure 2 (stimulus space in two dimensions), the primary purpose of the study was to generate a preference space in a third or z-dimension along with a preference surface (x,y,z) in the three-dimensional space. For each luminance-MTF combination, then, a preference must be estimated. A traditional means of estimating preference or scale values for stimuli in a multidimensional space is based upon unfolding models which attempt to satisfy the ordering of pairwise preferences by placing the points or stimuli in a multidimensional preference space. Relative distances from an ideal point in the space, typically the location of the observer in the space, represent preferences in the space.

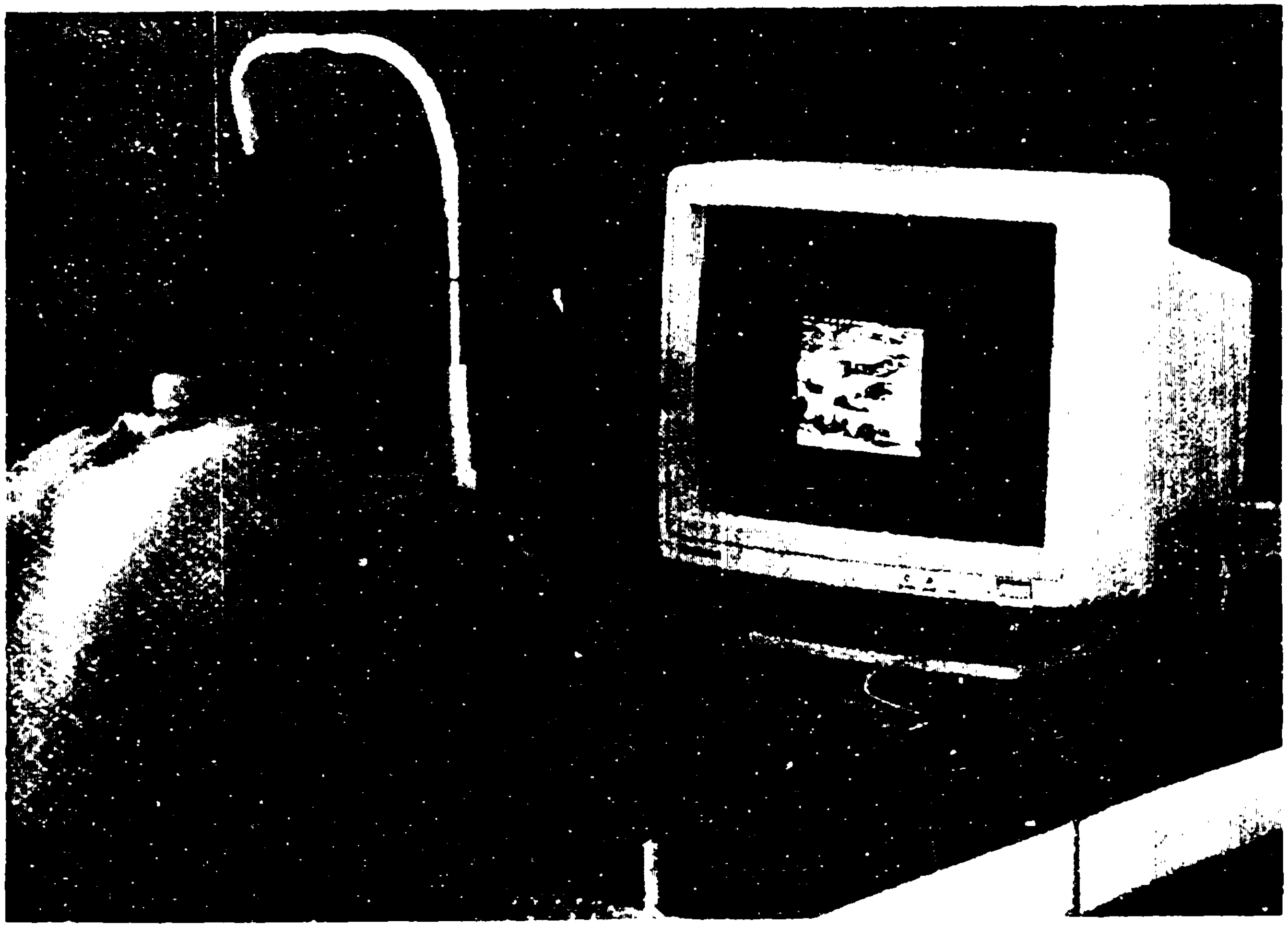


Figure 24
Experimental Apparatus Setup

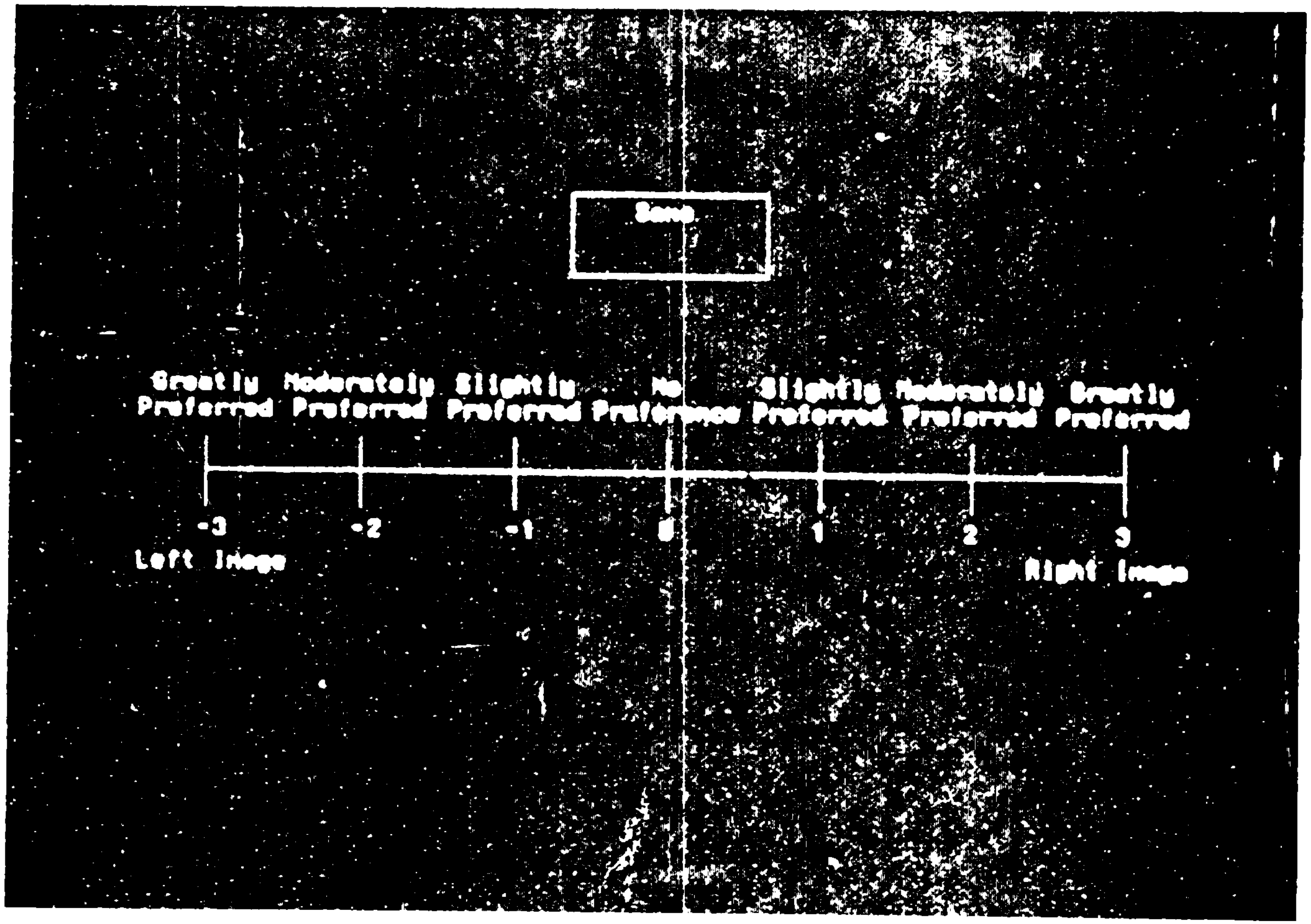


Figure 25
Response Rating Scale

In order to use an unfolding approach, the pairwise preference probabilities must be calculated. Preferences for the left or right image on individual trials are cumulated into empirical preference matrices (see the matrices in the appendix). The rows and columns of the matrices in the appendix denote the stimuli used in the paired comparison procedure. Each entry in the matrix, p(i,j), denotes the probability of a person preferring stimulus i to stimulus j. By symmetry p(i,j) = 1-p(j,i), and it was assumed that p(i,i), or the diagonal entries in the matrix, were .5 or 50%.

The appendix contains preference matrices for the present study. The first preference matrix is cumulated over all six observers and five images. Each cell or probability in this matrix is based upon 120 observations. It is also of interest to know whether individual observers differ in their preferences (i.e., the third component in Figure 3 [image-display-observer]) and whether the preferences change based upon the image (i.e., the first component in Fig. 3). In order to test these hypotheses, preference matrices were generated for individual observers combined over the five images and for individual images combined over the six observers. The second through sixth matrices in the appendix contain the preference matrices for each of the six observers combined over the five images. The probabilities in each cell for these matrices are based upon 24 observations. The final five matrices in the appendix contain preference matrices for each of the five images combined over the six observers. Probabilities in each cell for these matrices were based upon 20 observations.

Preference matrices based upon only a single observer and single image were also generated. Each probability in these matrices was based upon only four observations. These probability estimates were determined to be too unstable for use in modeling. The next section describes the modeling employed to unfold or generate the preference estimates for the MTF-luminance combination.

<u>The Bradley-Terry-Luce (BTL) Model of Choice</u>

The BTL Model (see De Soete and Carroll, 1992; or Bockenholt, 1992, for further introduction) is a psychological model of choice which can be fit to preference probabilities in a paired-comparison experiment. In the BTL Model, the probability that stimulus "i" is preferred to stimulus "j" is given as:

$$p_{ij}' = \frac{v(i)}{v(i) + v(j)}. \qquad (4)$$

v(i) and v(j) represent scale values or preference strengths for stimulus "i" and stimulus "j," respectively. In a paired-comparison experiment with N stimuli, the BTL model will have (N-1) strength or preference parameters to estimate. A single v(•) remains to serve as an anchor point. In the present study, v(20), the preference scale value for the 20th MTF-luminance combination (the worst MTF and the brightest luminance), was set to a value of unity and the remaining strength parameters were estimated for each stimulus or MTF-luminance combination relative to this anchor point.

The BTL model can be shown to be conceptually similar to the more popular model of choice, a Thurstone Case V Model, or a signal detection model with variances set to unity. By making a monotonic transformation, u(i)=ln(v(i)) or equivalently, $e^{u(i)}=v(i)$, Equation (4) can be transformed into:

$$p_{ij}'=1-\frac{e^{u(i)}}{e^{u(i)}+e^{u(j)}} = 1-\frac{1}{1+e^{-(u(i)-u(j))}} = 1-F[u(i)-u(j)]. \qquad (5)$$

F[•] in Equation (5) is the standard Logistic Distribution Function with two parameters, a and b, each set to unity. The mean of the Logistic Distribution in Equation (5) is one and the variance is approximately three ($\Pi^2/3$).

Figure 26 is a plot of the density function for the distribution in Equation (5). Note the similarity of this curve to the normal density curve. The abscissa denotes a random variable which is the difference between a variable representing stimulus "i" and one representing stimulus "j." This conceptual representation is similar to that of signal detection. On any trial when stimulus "i" and "j" are paired together, a preference or strength for each image is denoted by single random variable and a response is generated by comparing the two random variables and choosing the larger. The difference of the two random variables is distributed according to Figure 26.

In order to estimate the scale values, v(•), for a single empirical preference matrix, a computer routine was used to minimize the difference between the empirical preference matrix and one generated using Equation (4). Initial estimates of v(•) = 1 were used to generate an initial predictive matrix with p_{ij}' = .5. Through iteration and estimate of partial derivatives, the computer routine sequentially modified v(•) until it minimized the chi-square difference between the two matrices:

$$\chi^2 = \sum_{i,j} \frac{n(p_{ij}'-p_{ij})^2}{p_{ij}'} \qquad (6)$$

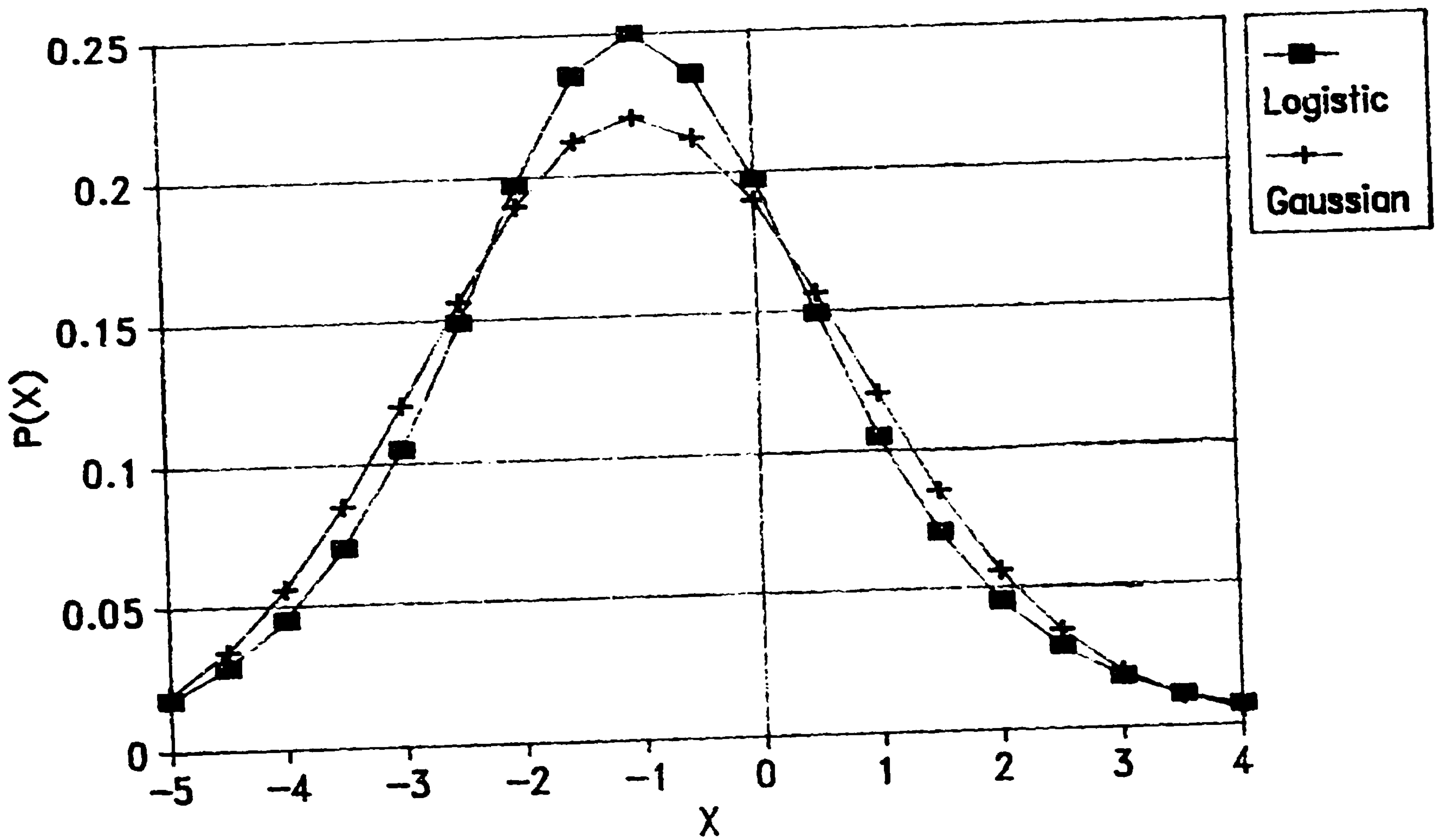


Figure 26
Logistic Density Function

where p_{ij}' denotes preference probabilities from the BTL model and p_{ij} are the empirically measured preference probabilities. Table 1 provides a listing of all the chi-square fits performed. The statistical test in each case is based upon the null hypothesis that the model provides a reasonable fit versus the alternative hypothesis that the model is rejected. The 20 x 20 matrix has $(20^2-20)/2 = 190$ independent data points to fit since $p_{ii} = .5$ and $p_{ij} = p_{ji}$. The BTL model has 20-1 = 19 independent parameters for this study. The degrees of freedom (df) for the chi-square fit would then be 190-19 = 171. For $\alpha = .2$ and df = 171, the critical $X^2 = 186.35$, so none of the model fits shown in Table 1 could be rejected.

Table 1

Chi-Square Fit of The Bradley-Terry-Luce Model to Empirical Data

Data matrix	Observed chi-square
average over all observers & images	19.03
observer BA averaged over all images	7.59
observer EU averaged over all images	11.90
observer JB averaged over all images	19.61
observer KK averaged over all images	30.79
observer MB averaged over all images	15.83
observer MD averaged over all images	16.60
airport image averaged over all observers	9.52
crop image averaged over all observers	14.07
mountain image averaged over all observers	6.66
ocean image averaged over all observers	13.00
pine image averaged over all observers	14.57

Although the BTL model fit could not be rejected, some models are so pervasive that they will fit any set of data. To test this hypothesis, the BTL model was fit to a 20 x 20 matrix containing probabilities generated using a uniform random number generator. These matrices were constrained in the same fashion as the empirical preference matrices so that $p_{ij} = p_{ji}$ and $p_{ii} = .5$. The X^2 fit to sets of randomly generated matrices (using n = 20,24, and 120 as shown in Equation (6)) was consistently greater than 600, large enough to be rejected at $\alpha = .01$ levels.

The scale values, v(•), generated by the BTL model fit for the empirical matrices are shown in Table 2. In Table 2, there are 20 scale values in each column or for each condition. Each scale

Table 2. BTL Scale Values for All Experimental Conditions

(1) MTF	(2) LUM	(3) ALL	(4) BA_ALL	(5) EU_ALL	(6) JB_ALL	(7) KK_ALL	(8) MB_ALL	(9) MD_ALL	(10) A_ALL	(11) C_ALL	(12) M_ALL	(13) O_ALL	(14) P_ALL
1	1	40.023	1363.191	22.881	30.043	51.235	9.836	384.816	142.261	36.656	283.495	23.981	18.337
1	2	137.594	7039.538	150.458	116.083	171.424	26.263	802.603	442.524	169.196	1271.866	66.974	63.014
1	3	256.583	25372.46	559.019	382.404	252.756	27.129	930.963	1007.952	556.392	3191.456	116.174	52.604
1	4	227.986	69911.03	761.11	449.527	192.117	13.624	620.182	822.751	893.558	3448.076	91.409	29.856
2	1	13.447	126.118	8.353	10.506	21.289	4.096	90.896	41.583	9.074	50.153	8.313	8.822
2	2	46.102	542.666	56.906	52.981	74.075	11.187	215.693	158.517	58.303	223.457	23.216	24.951
2	3	69.751	1656.255	138.364	107.359	120.048	9.786	209.385	212.272	175.216	466.741	35.276	18.609
2	4	67.743	2013.66	272.658	106.69	71.74	7.538	148.941	220.018	223.405	553.067	30.904	12.544
3	1	3.738	12.616	2.14	2.732	7.021	1.99	14.072	9.629	1.877	11.018	2.481	3.14
3	2	10.336	38.584	11.517	9.366	23.273	3.547	29.518	24.925	10.127	33.926	6.038	7.092
3	3	16.792	97.484	39.836	19.084	27.23	3.938	35.321	38.79	26.394	61.701	10.183	7.838
3	4	15.271	189.676	41.366	18.324	18.056	2.629	25.481	30.183	44.579	56.004	9.469	4.512
4	1	1.08	1.39	0.487	0.896	1.93	1.01	2.888	1.722	0.649	1.774	0.97	1.122
4	2	2.547	4.421	2.704	2.299	5.323	1.279	5.88	4.301	1.719	4.913	1.796	2.697
4	3	3.646	11.601	5.842	3.314	5.921	1.392	6.201	6.443	4.342	5.443	3.059	2.713
4	4	3.39	12.781	8.649	2.95	4.041	1.153	5.242	5.372	5.982	5.607	2.565	1.875
5	1	0.225	0.199	0.05	0.398	0.052	0.549	0.436	0.187	0.189	0.233	0.185	0.364
5	2	0.359	0.307	0.185	0.492	0.191	0.534	0.541	0.32	0.275	0.396	0.307	0.533
5	3	0.489	0.435	0.435	0.553	0.465	0.505	0.524	0.465	0.468	0.445	0.47	0.618
5	4	1	1	1	1	1	1	1	1	1	1	1	1

(1) MTF Area - 1 = 6.18
2 = 4.73
3 = 3.79
4 = 3.15
5 = 2.34

(2) Luminance - 1 = 4.0 fL
2 = 6.3 fL
3 = 8.5 fL
4 = 11.1 fL

(3) BTL scale values averaged over all observers and images
(4) BTL scale values for observer BA averaged over all images
(5) BTL scale values for observer EU averaged over all images
(6) BTL scale values for observer JB averaged over all images
(7) BTL scale values for observer KK averaged over all images
(8) BTL scale values for observer MB averaged over all images
(9) BTL scale values for observer MD averaged over all images
(10) BTL scale values for Airport Image averaged over all observers
(11) BTL scale values for Crop Image averaged over all observers
(12) BTL scale values for Mountain Image averaged over all observers
(13) BTL scale values for Ocean Image averaged over all observers
(14) BTL scale values for Pines Image averaged over all observers

value represents the preference for an MTF-luminance combination relative to the other 19 MTF-luminance combinations. These scale values complete the three-dimensional preference space. Hypothetically, a display would be represented by its MTF-luminance combination or (x,y) in the plane of the three-dimensional space and the predicted scale value (z-axis) would locate the display in (x,y,z).

From this study, 20 points (5 MTFs x 4 luminance levels) were generated in the three-dimensional space for each of the 12 conditions shown in Tables 2 and 3. Figure 27A shows the 20 BTL scale values for the preference data averaged over observers and images. Figure 27B is a three-dimensional plot of the 20 points where a smoothing algorithm was used to generate the curve drawn through the points from Figure 27A.

A goal of this research was to be able to compare the MTF-luminance combinations from two displays and predict which display would be preferred. From the curve in Figure 27B, it would be difficult to locate the preference scale value in the z-dimension based upon an MTF-luminance combination. A predictive equation would be useful for this purpose. In the following section, regression is used to generate surface equations for Figure 27B.

Generation of Predictive Equations

Statistical regression software was used to generate prediction equations for the BTL scale values as a function of MTF and luminance. Because the range of scale values in Table 2 was so large, a logarithmic transformation of the data was used. The form of the predictive equations was:

$$\log_{10} (\text{scale value} + 1) = f(\text{MTF, Luminance}) \qquad (7)$$

where the goal was to estimate the form of the function f. From fits of linear, quadratic, and cubic polynomials, quadratic formulas were chosen based upon their R^2 values. Table 3 shows the coefficients and R^2 fits for the 12 matrix conditions based upon the following equation:

$$\log_{10} (\text{scale value} + 1) = a + b(\text{MTF}) + c(\text{Lum}) + d(\text{MTF})(\text{Lum}) + e(\text{MTF})^2 + f(\text{Lum})^2. \qquad (8A)$$

For data averaged over all observers and images, the prediction of scale values is as follows:

$$\textit{scale value} = 10^{-2.51+.778(MTF)+.226(LUM)-.047(MTF)^2-.016(LUM)^2+.020(MTF)(LUM)} - 1 \qquad (8B)$$

Table 3

Quadratic Regression Coefficients and R^2 Values to the Equation:

$$\log_{10}(\text{scale value} + 1) = a + b(\text{MTF}) + c(\text{Lum}) + d(\text{MTF})^2 + e(\text{Lum})^2 + f(\text{MTF})(\text{Lum}).$$

Preference condition	Intercept a	MTF b	LUM c	MTF^2 d	LUM^2 e	MTFXLUM f	R^2
average	-2.51	.778	.226	-.047	-.016	.020	.983
observer BA	-3.18	1.033	.131*	-.042	-.013*	.050*	.991
observer EU	-3.15	.936	.253	-.080	-.018	.044	.981
observer JB	-2.16	.610	.197*	-.042	-.017*	.040	.974
observer KK	-3.78	1.318	.337	-.102*	-.022	.013	.985
observer MB	-1.24	.300*	.200	-.003*	-.013*	.002	.951
observer MD	-3.55	1.367	.243	-.079*	-.015	.001	.988
airport image	-3.55	1.266	.253	-.087	-.018	.021	.986
crop image	-2.57	.750	.181*	-.054	-.014	.043	.976
mountain image	-3.40	1.163	.224	-.073	-.018	.036	.985
ocean image	-2.07	.596	.211	-.030	-.014	.014	.983
pine image	-2.20	.619	.306	-.029	-.019*	**	.971

* $P > .05$ (not significantly different from zero at $\alpha = .05$)

** $< 10^{-3}$ and $P > .05$

The preference conditions in Table 3 have many commonalities with respect to their regression coefficients. For example, in most instances the linear slope attributed to MTF is two to three times the slope of luminance. The relationship across the two dimensions (MTF and luminance) and their effect on preference was the most important effect of interest in this study. Within the MTF and luminance bounds of this experiment, a change in the area under the MTF curve has more of an effect on preference than an equivalent numerical change in luminance. It is critical, though, that this statement not be misinterpreted, as it cannot be used to mean that changes in MTF are more important than changes in Luminance to viewer preference. Suppose, for example, that Luminance had been measured in tenths of a FootLambert instead of FootLamberts. Then, the results here would show that the change in Luminance had a much greater effect on viewer preference than the equivalent numerical change in MTF. Thus, across the two dimensions, we can only make statements such as a change of 1 unit of MTF (percent contrast X cycles/degree of visual angle) is equivalent to a change of 2 units of Luminance (FootLamberts) with respect to viewer preference. A numerical example given below will illustrate the trade-off across these two dimensions.

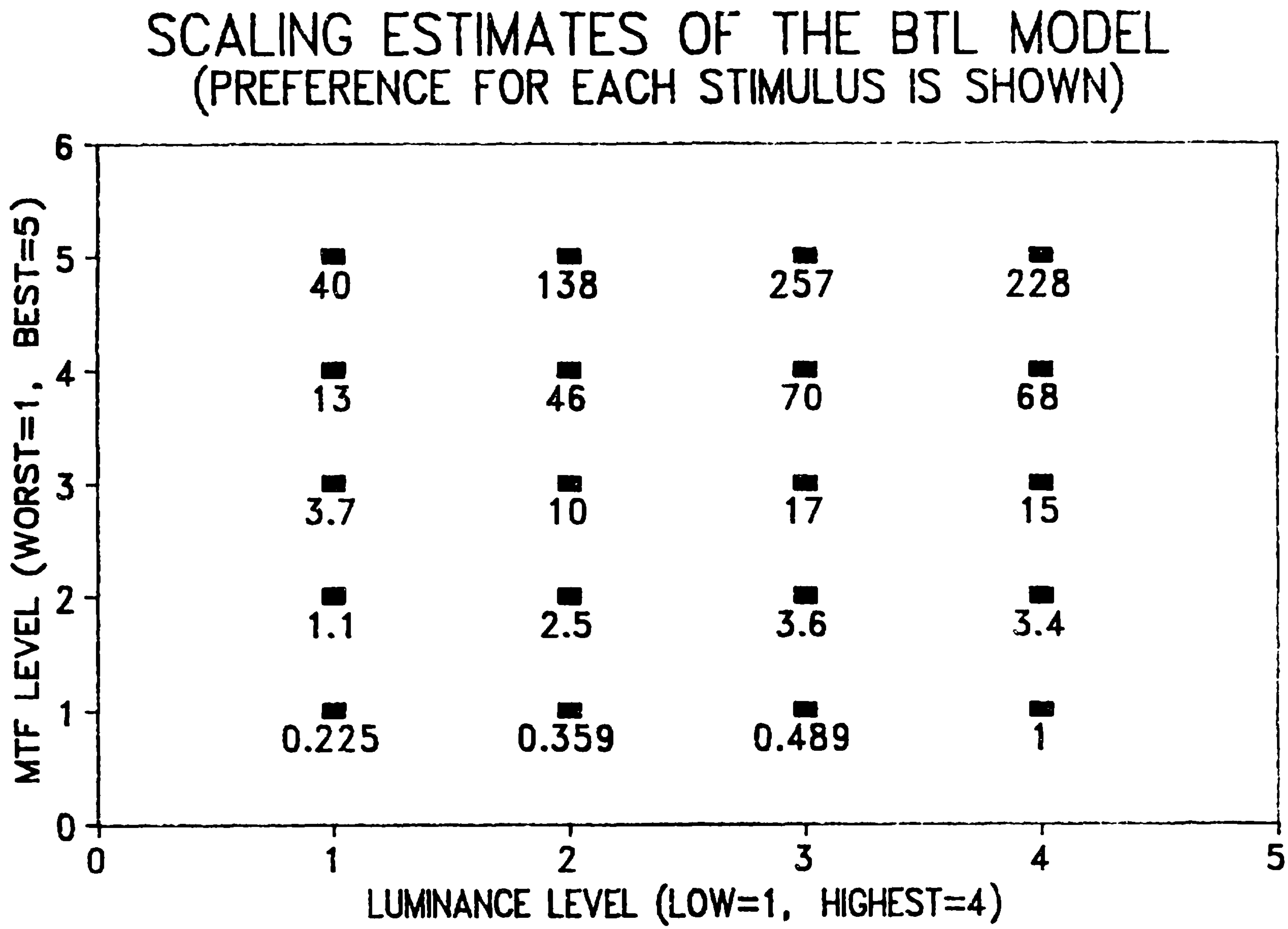


Figure 27A
BTL Estimates for Preference Data Averaged Over Observers and Images

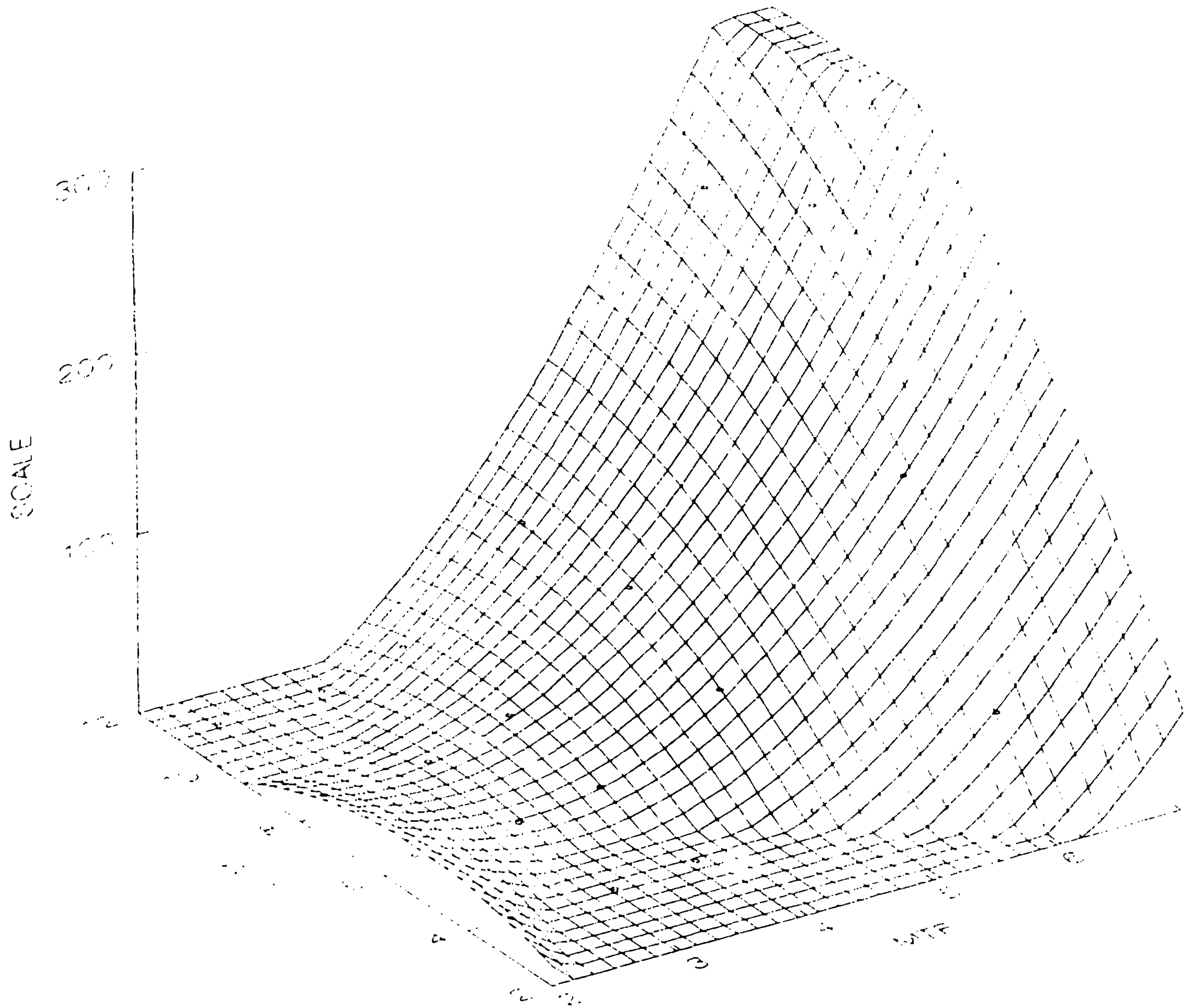


Figure 27B
Three-Dimensional Plot of Preference Scale Values

The linear effect (i.e., the parameter c) due to luminance was also statistically insignificant in three instances as noted by asterisks in Table 3. Figures 28A-E show two-dimensional cross sections of preference as a function of luminance for the five levels of MTF for each of the six observers. Under the optimal level of MTF (Fig. 28A), preference is nonmonotonic as a function of luminance for three of the six observers. For three of the observers, preference first increases and then decreases as luminance level increases. More detailed experimental work would be required to find the asymptotic level of luminance. Most likely, though, the asymptotic level of luminance depends upon the background and pre/post-trial luminance. In this experiment, the pre/post-trial luminance was quite dark (i.e., < .1 fL) and the background or display luminance surrounding the images was approximately 7 fL. As a result, the most preferred luminance was probably in the range of 8-11 fL, much lower than if a bright background had been employed.

Coefficients for squared terms in the regression equations were consistently negative, denoting a slowing down or curvilinear effect of the $\log_{10}$ of preference as one or the other of MTF and luminance increased. This does not necessarily mean that preference tended to asymptote or actually decrease as MTF or luminance independently increased. It implies that, within the ranges of MTF and luminance studied, the rate of increase in the log of preference was negatively accelerated or slowing down.

Finally, the interaction coefficients (f) in Table 3 were all positive. A likely explanation for this finding is that preference broke down at low levels of either MTF or luminance (see Figs. 28D and E). Note that the scaling on the y-axis in Figures 28D and E is much less than the scaling in Figures 28A-C. For the two most blurred MTFs, the amount of luminance mattered less because the effect of the MTF dominated the observer's perception of the image. Similarly, the lack of image quality at the lowest luminance level also dominated the observers' preference response.

As an example of the use of the Equations in Table 3, consider the comparison of two displays, A and B. Using Equation (8) and the regression coefficients from Table 3, Displays A and B may be compared to one another if they are characterized by their MTF-luminance combinations. Assume the two displays have the following characteristics with respect to their MTF areas and average luminances (in fL):

$MTF_A = 4.$ $\qquad$ $MTF_B = 5.$
$Lum_A = 8.$ $\qquad$ $Lum_B = 5.$

This case exemplifies the trade-off problem across dimensions where display A has a lower or worse MTF but also has more luminance. By

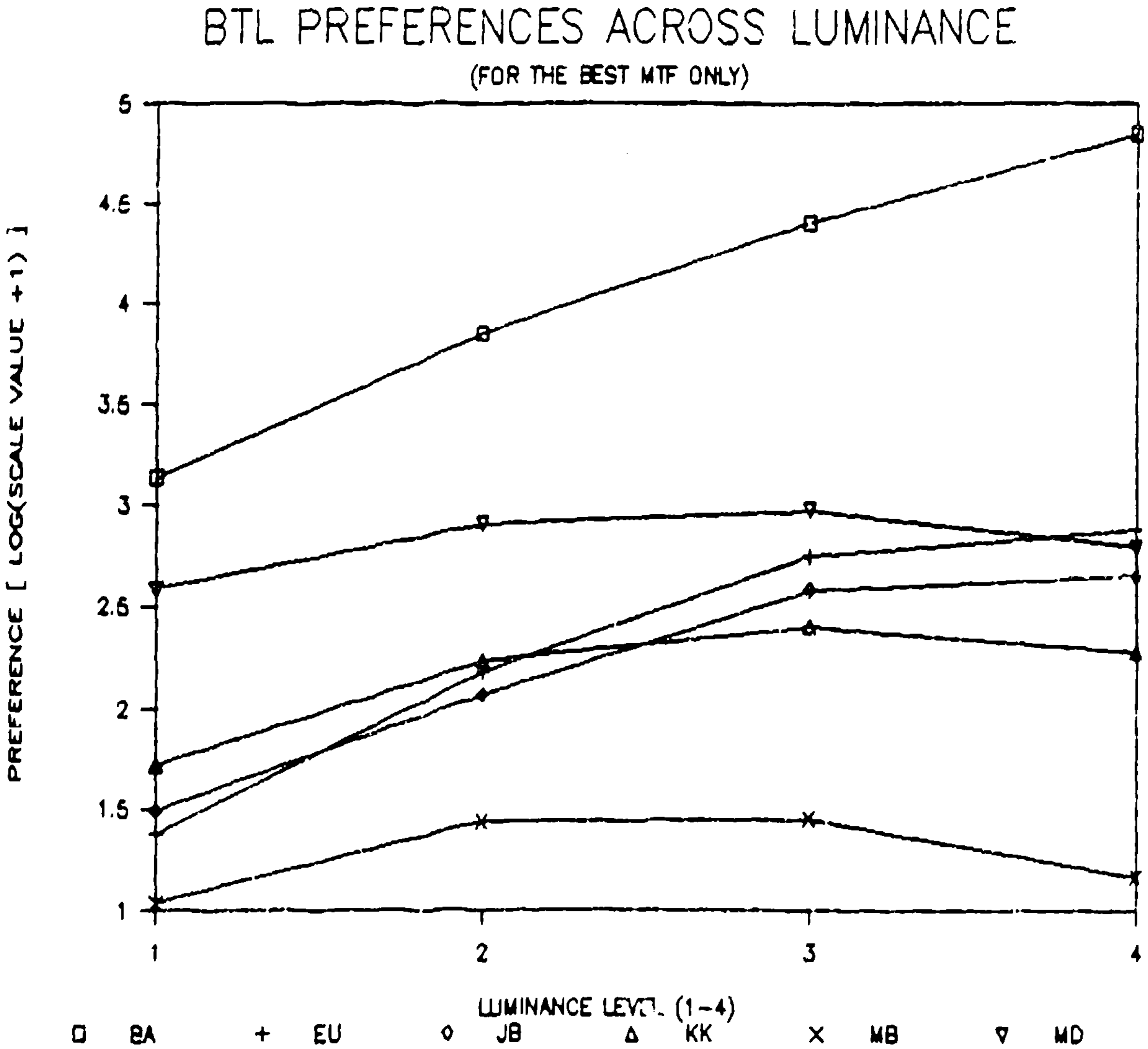


Figure 28A
Preference as a Function of Average Display Luminance
(MTF Area = 6.18)

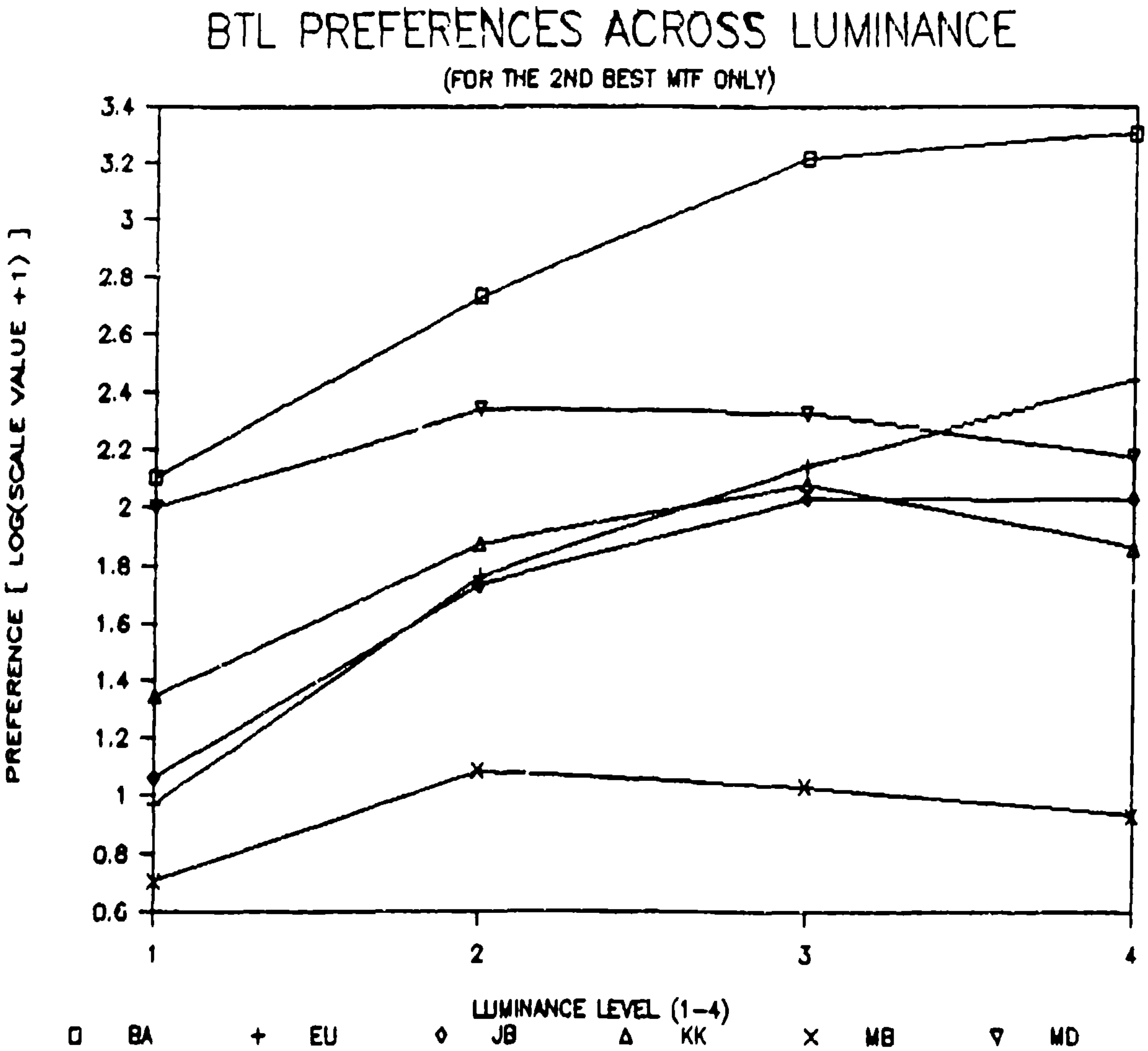


Figure 28B
Preference as a Function of Average Display Luminance
(MTF Area = 4.73)

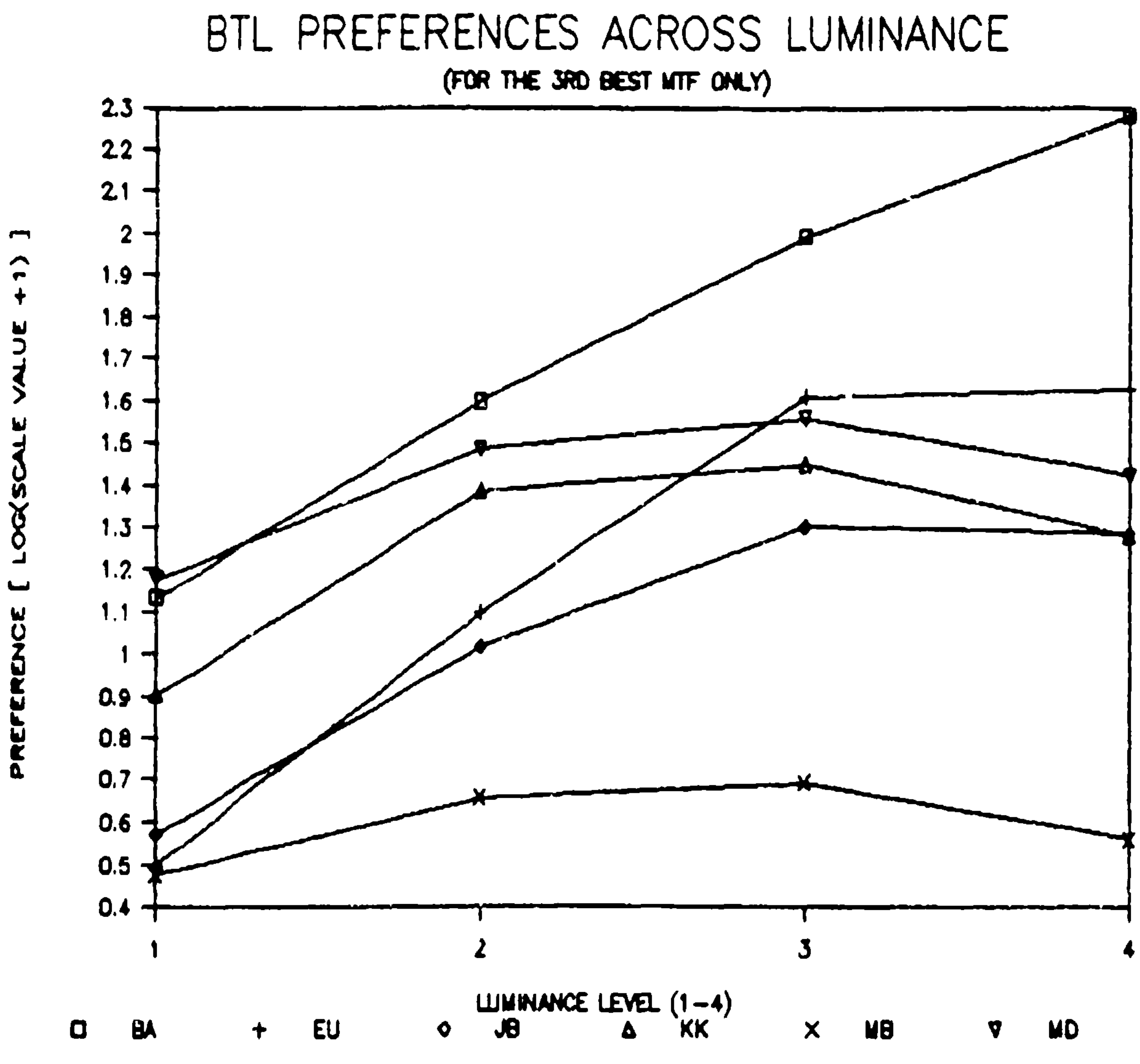


Figure 28C
Preference as a Function of Average Display Luminance
(MTF Area = 3.79)

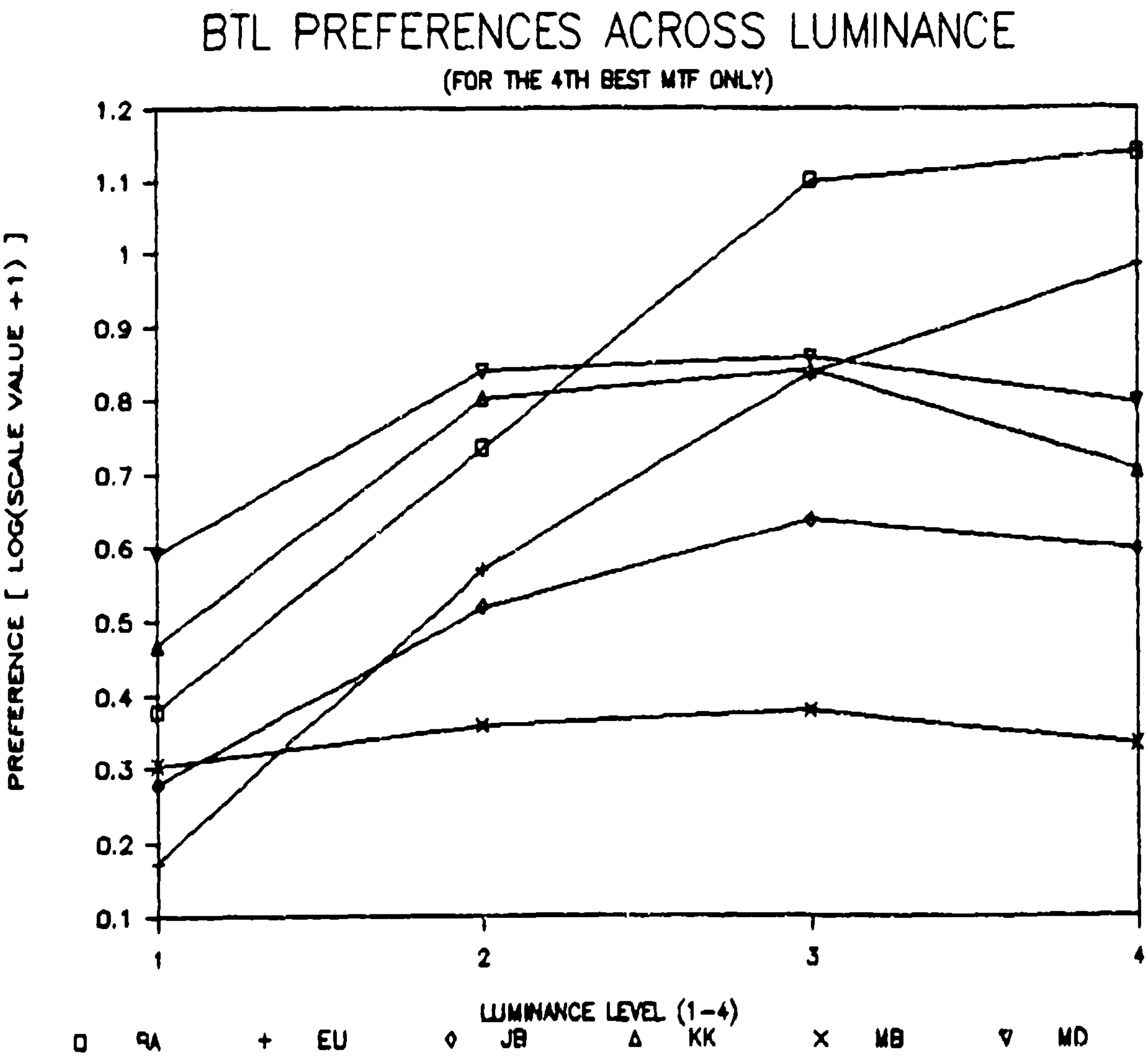


Figure 28D
Preference as a Function of Average Display Luminance
(MTF Area = 3.15)

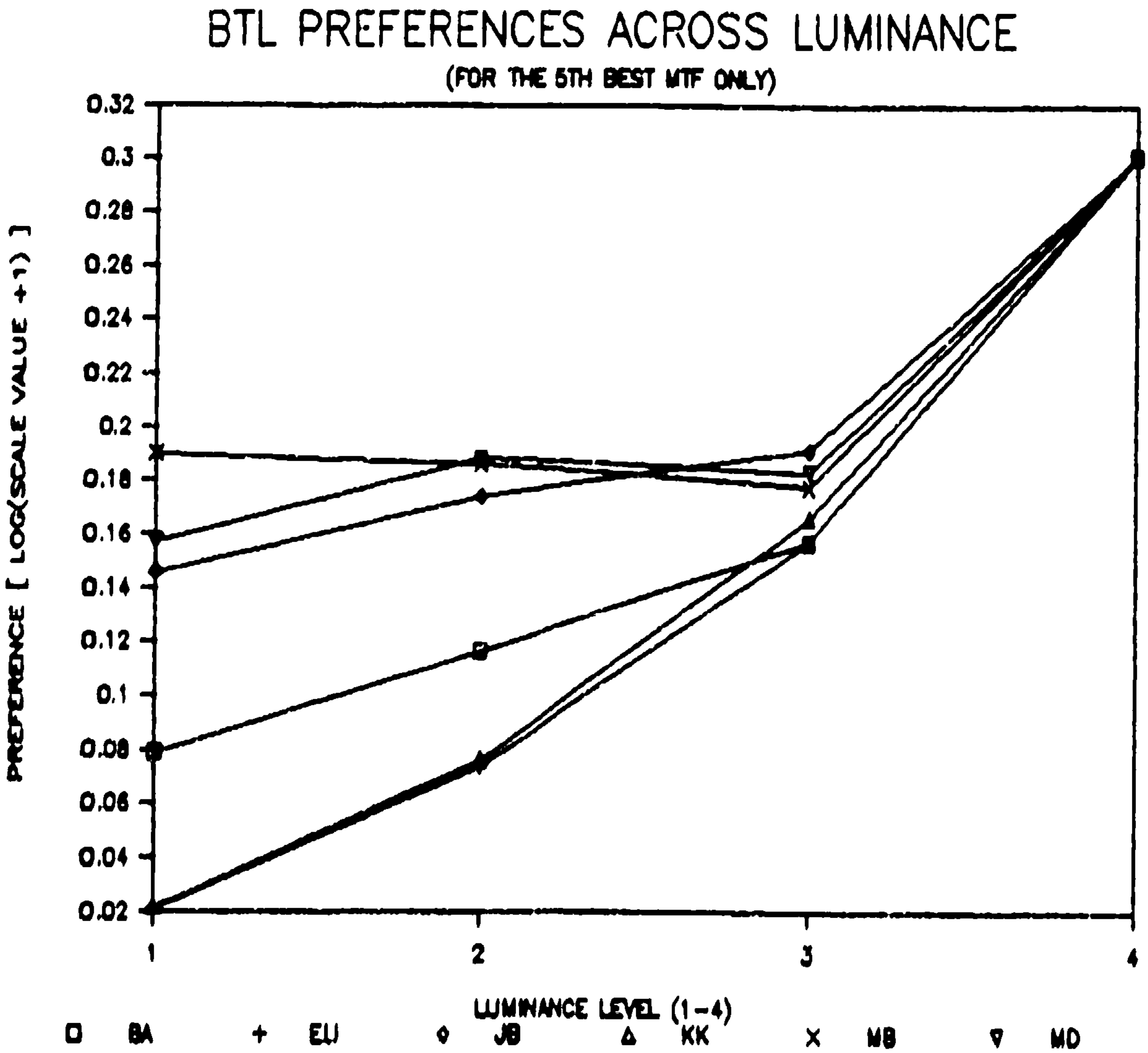


Figure 28E
Preference as a Function of Average Display Luminance
(MTF Area = 2.34)

solving Equation (8) using the coefficients from the regression to the overall average in Table 3, the result obtained yields v(A) = 19.3 and v(B) = 27.8. Since the scale value for display B is greater than that of display A, the conclusion would be that display B is preferred to display A.

With most image quality metrics, the comparison of the two displays would either end here or anecdotal evidence would be used in concluding whether this difference was significant. An advantage of this empirically generated prediction, though, is that the scale values can be interpreted through the experimental paradigm. Applying the scale values v(A) = 19.3 and v(B) = 27.8 to Equation (4) yields:

$$p_{AB}' = \frac{v(A)}{v(A) + v(B)} = \frac{19.3}{19.3 + 27.8} = .41 \qquad (9)$$

or the probability that display A is chosen over display B is estimated as 41%. The data used to generate the scale values were cumulated over observers and images. Generalizing over these two factors, it may be said that if a number of observers viewed a number of images on display A and display B, they would prefer the images from display A on approximately 41% of the trials. It may be noted here that the BTL preferences given here are forced-choice preferences. It is assumed now that p_{BA}, the probability that B is preferred to A, is 59% and that no indifferent trials exist.

In order to directly compare two displays based upon their MTF-luminance combinations, Equation (8) must be used with the coefficients from Table 3. Equation (8) can be plotted as a surface as in Figure 27B but it is difficult, if not impossible, to visually locate the preference values for MTF-luminance combinations.

Two displays may be compared indirectly in a two-dimensional graph by comparing both to a reference display device. Figure 29 is one example of this, showing a two-dimensional graph with preference probability on the y-axis, MTF area on the x-axis, and multiple curves generated which correspond to a range of luminance levels. Using the previous example with two displays, A and B, the approximate preference probability of Display A over the standard is .45 and the approximate preference probability of Display B over the reference is .58. Because .58 > .45, display B would be preferred to Display A, although the exact preference probability of display B over display A could be computed only by calculating the actual BTL scale values of both displays and using the scale values in Equation (4).

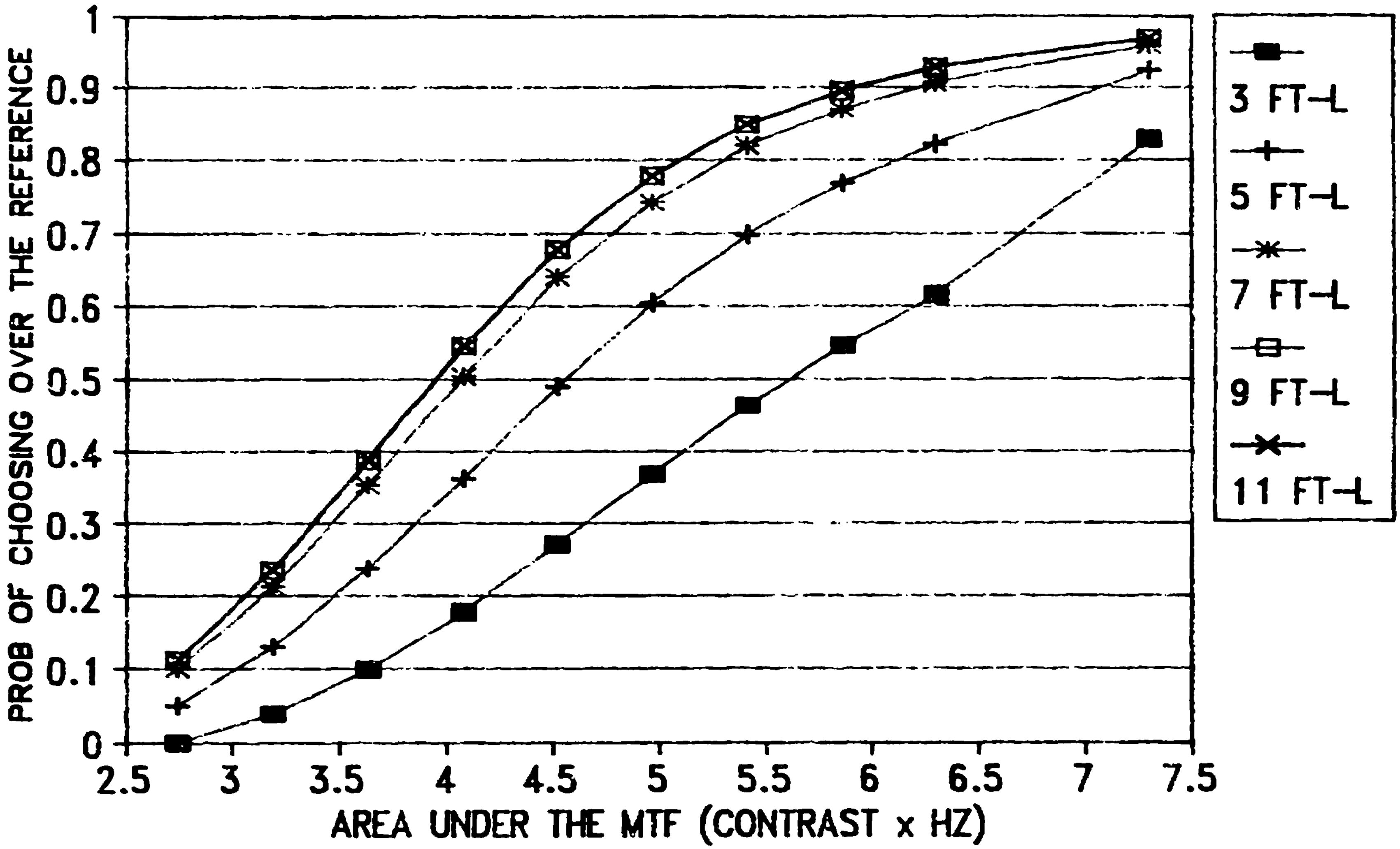


Figure 29
Two-Dimensional Graph Predicting Preference with Respect to a Reference Display

The predictive validity of the graphs and equations discussed in this section apply only within the ranges of MTF area (in Michelson Contrast x cycles/degree of visual angle) and luminance (in fL). The levels of MTF area used in the present study were between 2.34 and 6.18. From pilot work performed previously (Evans, 1993), differences between the MTF areas become less discriminable as magnitude grows. Therefore, extrapolation of the predictive equations beyond an MTF area of 6.18 would have little validity.

Average luminance levels of the images in this study ranged from 4.0 to 11.1 fL. Although preference for the average luminance level peaked somewhere below 11.1 fL, the ambient illumination (<.1 fL) and CRT luminance surrounding the images (approximately 7 fL) are likely to be moderating factors in this finding.

Another problem in the use and generality of the predictive equations is a logical problem. When using a display, if the display is capable of putting out more luminance, it is logical that the display could also produce lower luminance levels. For example, if two displays are equivalent in the MTF dimension but one display is capable of generating more luminance, it is not logical that the display with less luminance capabilities be preferred. Because of this fact, the luminance value used in predicting preference should be the luminance less than or equal to the average luminance level of the display which maximizes preference. The scale value, then, should be:

$$\text{MAX } \{v\} \text{ over } \{\text{MTF}, L \leq \text{LUM}\} \tag{10}$$

or that we choose the maximum preference scale value for the MTF area of the display and all luminances less than or equal to the average luminance of the display. Using the quadratic prediction in Equation (8), v can be maximized with respect to luminance by taking the partial derivative of the function with respect to luminance and setting it equal to zero. The result yields:

$$L_{max} = -(c + f(\text{MTF}))/(2e) \tag{11}$$

where MTF is the MTF area for the particular display, Lmax is the luminance level which maximizes the preference, v, for the MTF given, and $c = .226$, $e = -.016$, and $f = .02$ are coefficients taken from the averaged regression in Table 3. Using the previous example where $\text{MTF}_A = 4$, $\text{LUM}_A = 8$, $\text{MTF}_B = 5$, and $\text{LUM}_B = 5$, Lmax for Displays A and B would be 9.6 and 10.2 fL, respectively. If the average luminance of display A or B was greater than their Lmax value, the luminance value of that display would be replaced by Lmax for the purpose of computing preference scale values. Figure 30 is a graph of Equation (11) for a range of MTF areas used in the study. Prior to using Equation (8) or the graph in Figure 29 for

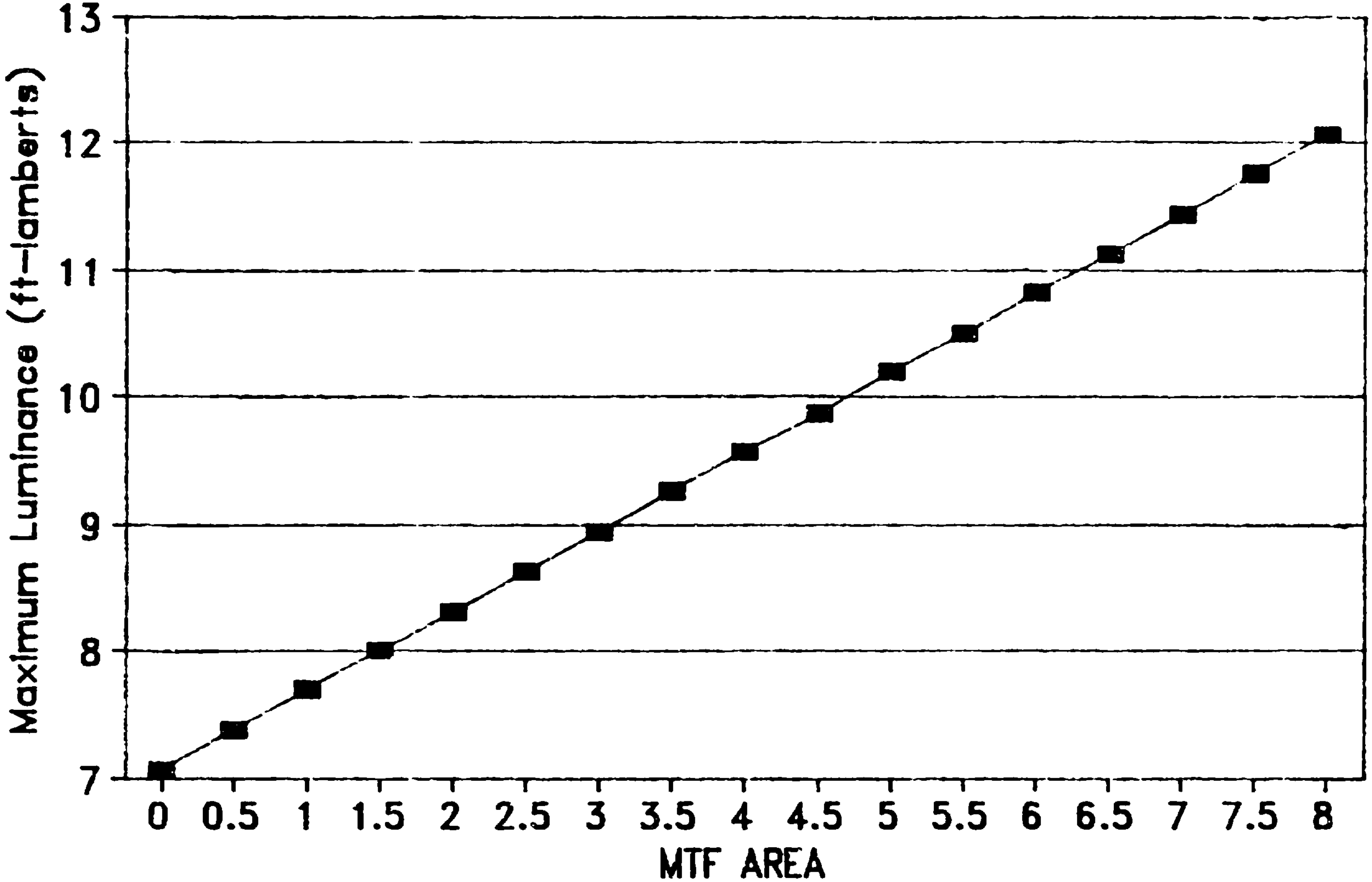


Figure 30
Prediction of Luminance which Maximizes Preference as a Function of MTF Area

predicting preference, then, it would be necessary to use Figure 30 in order to find Lmax first. As mentioned, it is suspected that the predictions obtained in this section are dependent upon a number of moderating factors (e.g., background luminance, field of view). Possibly, one of the most important moderating factors is the set of imagery used in the study. Although there are no rigorous or well-accepted methods for analyzing the information content in imagery, simple observation of the images in Figures 7 through 11 would lead one to conclude that the Ocean Scene (Fig. 10) has less information than the other images with which to develop preference standards. In addition, the Airport Scene (Fig. 7) has a number of man-made features which might hypothetically help in discriminating MTF-luminance combinations. In the next section, statistical analysis of the effect of the images on preference is presented.

The Effect of Image on Preference

As emphasized previously, none of the results obtained in this study should be considered as being independent of the imagery being viewed. Referring back to Figure 3, the information contained in the original image can affect processes in the display and observer subsystems. The five images (airport, crop, mountain, ocean, and pines) chosen for viewing in this study are visually quite distinct. They were chosen as stimuli in this study because, in a multidimensional scaling study performed by Kleiss (1992), these stimuli were quite distinct from one another in a two-dimensional similarity mapping of natural images. It was desirable in this study to determine if preference for MTF-luminance combinations might depend upon the image type (or if preferences were consistent across a diversity of image types).

An analysis of covariance was performed on the BTL scale values obtained for the five images in Table 2. Note that the BTL scale values in Table 2 pertaining to the five images (airport, crop, mountain, ocean, and pines) were averaged over observers. The averaging was performed in order to increase the number of observations upon which the preference probabilities were based. In order that the scale values for each of the five data sets in Table 2 meet the homogeneity of variance assumption, the analysis was performed on a logarithmic transformation of the scale values.

The average scale values for the five images were as follows: airport (159), crops (111), mountain (484), ocean (22), pines (13). Recall that the scale values are used in Equation (4) to obtain the preference probabilities. Higher scale values may be interpreted as polarization of preference probabilities towards zero and one. This, in turn, may be interpreted as an increased preference discrimination across the MTF-luminance combinations. From the five means, it can be seen that there was substantially less discrimination with the ocean and pines images than the other three. It was suspected a priori that the ocean image would not provide as much discriminatory power when viewing the MTF-luminance combinations, but the low discriminability of the pines image was unexpected. However, the DAC distribution of the pines image (see Fig. 4) proved to be more bimodal than the other images, divided between the brightness of the sky at the top of the image and the remainder of the image being underexposed (in the original image). The sky provided little information and the remainder of the image was too underexposed to provide much discrimination in the preference response. Although it was expected that the airport image with its artificial components would yield the highest amount of discriminability, the mountain image with its speckled pattern of desert terrain provided the highest level of discrimination.

The analysis of covariance is shown in Table 4. The factors MTF and LUMINANCE in Table 4 denote linear regressions for the two factors, each based upon one degree of freedom. The images were shown to be significantly different at $p<.0001$, providing evidence that the image can significantly affect the perception of the MTF-Luminance filters. The two interaction terms, MTF x IMAGE and LUM x IMAGE, were both significant at $p<.001$. The interpretation of these results is that the slope of predicting the dependent variable ($\log_{10}$ one plus the scale value) as a function of MTF or luminance changes significantly across images. Table 5 shows the slopes for predicting the dependent variable ($\log_{10}$ one plus the scale value) as a function of either MTF or luminance for each of the five images.

As with the linear coefficients from the quadratic regression in Table 3, the slopes are all positive and changes in MTF area have a larger effect on the dependent variable than changes in luminance.

Table 4

Analysis of Covariance of BTL Scale Values for Determining Image Differences

Source	DF	SS	MS	F-Value	Significance
MTF	1	59.08	59.08	1175.14	<.0001
Luminance	1	3.59	3.59	71.32	<.0001
Image	4	7.73	1.93	38.42	<.0001
MTF x Image	4	4.37	1.09	21.75	<.0001
LUM X Image	4	1.08	.27	5.36	.0007
Error	85	4.27	.05		
Total	99	80.12			

Table 5

Regression Slopes for Dependent Measure ($\log_{10}$ one plus the scale value) as a Function of MTF or Luminance

	Independent Variable	
Image	MTF	Luminance
Airport	.525	.030
Crops	.467	.059
Mountains	.621	.040
Ocean	.343	.025
Pines	.284	.006

CONCLUSIONS

The work in this study represents an empirical approach in modeling multidimensional aspects of image/display quality. Within the ranges of MTF area (2.34 - 6.18 in % Michelson Contrast X cycles/degree of visual angle) and average display luminance (4.0 - 11.1 fL) used in the study, the BTL Model of Choice and regression were used to predict display preference as a function of MTF area and average luminance. The results obtained from the modeling allowed a quantitative comparison across the two dimensions of interest. Within the ranges of MTF area and average display luminance studied, the following general results were noted:

(1) Changes in MTF area (% Michelson Contrast X cycles/degree of visual angle) had a much greater effect on observer preference than did changes in average display luminance (in fL).

(2) Changes in viewer preference were monotonically related to changes in MTF area.

(3) Changes in viewer preference were monotonically related to changes in luminance (FL) for three observers. For the remaining three observers, maximum preference occurred at the second brightest luminance level.

A more quantitative comparison of the relative effect of the two factors on preference was obtained using a quadratic regression.

Shortcomings of the multidimensional approach used here consisted of (a) the large number of observations required, (b) the limits on the ranges of factors studied, and (c) the difficulty in manipulating display parameters.

With respect to (a) in the present study, there were 20 MTF-luminance combinations and the paired-comparisons approach required a total of 400 possible stimulus combinations. To generate preference probabilities, each of these 400 possible combinations had to be presented multiple times, enough to generate stable probability estimates.

With respect to (b), limits on the ranges of factors studied, the modeling approach used in this study was predicated on the use of a matrix of paired-comparison preference probabilities. To unfold these preferences into a three-dimensional space, all of the stimuli must be related directly or indirectly to all other stimuli. If any stimulus or set of stimuli is not related to other stimuli (i.e., all preference probabilities are either zero or one), then this stimulus or set of stimuli cannot be placed into the multidimensional space. This restriction oftentimes limits the range of stimuli which may be studied. In the case of the present study, the range of luminance values and the range of MTF areas was restricted.

With respect to (c), the difficulty in manipulating display parameters, it was necessary to compress the original images in order to manipulate average luminance. In addition, the display used in presenting the stimuli restricted the upper limit of MTF areas that could be generated. The technique used in this study of using a single display to present all of the imagery allowed control of many extraneous factors that would come into play if multiple displays were used to present imagery. The limitation of this technique is the restriction in the range of the variables studied.

Finally, this study represents an attempt at manipulating multiple display factors toward the purpose of comparing their relative effects on image quality. In order to compare the quality provided by practical display systems, it will be necessary to quantitatively compare trade-offs across dimensions simply because, in many instances, design of these systems involves trade-offs of the display parameters.

REFERENCES

Barten, P.G.J. (1984). Spot size and current density distribution of CRTs. Proceedings of the SID, Vol. 25(3), 155-159.

Barten, P.G.J. (1989). The square root integral (SQRI): A new metric to describe the effect of various display parameters on perceived image quality. SPIE Vol. 1077 Human Vision, Visual Processing, and Digital Display, pp. 73-82.

Barten, P.G.J. (1990). Evaluation of subjective image quality with the square-root integral method. Journal of the Optical Society of America, 7(10), 2024-2031.

Barten, P.G.J. (1991). Evaluation of the effect of noise on subjective image quality. SPIE Vol. 1453 Human Vision, Visual Processing, and Digital Display II, pp. 2-15.

Bockenholt, U. (1992). Multivariate models of preference and choice. In F.G. Ashby (Ed.), Multidimensional Models of Perception and Cognition. Hillsdale, NJ: Lawrence Erlbaum Associates.

De Soete, G., & Carroll, J.D. (1992). Probabilistic multidimensional models of pairwise choice data. In F.G. Ashby (Ed.), Multidimensional Models of Perception and Cognition. Hillsdale, NJ: Lawrence Erlbaum Associates.

Evans, R.J. (1990). Image quality metrics and application of the square root integral (SQRI) metric: An overview. (AFHRL-TR-90-56, AD-A229 753). Williams Air Force Base, AZ: Operations Training Division, Air Force Human Resources Laboratory.

Evans, R.J. (1993). Image quality and the display modulation transfer function (MTF): Experimental findings. (AL/HR-TR-1993-0131). Mesa, AZ: Aircrew Training Research Division, Armstrong Laboratory.

Kleiss, J.A. (1992). Perceptual dimensions of visual scenes relevant for simulating low-altitude flight. (AL-TR-1992-0011). Williams Air Force Base, AZ: Human Resources Directorate, Aircrew Training Research Division.

Kusaka, H. (January, 1989). Consideration of vision and image quality - Psychological effects induced by picture sharpness. Paper presented at the SPIE/SPSE Conference, Los Angeles, CA.

Snyder, H. (1985). Image quality: Measures and visual performance. In L.E. Tannas, Jr., (Ed.), Flat Panel Displays and CRTs. NY: Van Nostrand Reinhold Company, Inc.

APPENDIX

PREFERENCE MATRICES

APPENDIX TABLE 1: PROBABILITY OF PREFERRING ROW STIMULUS OVER THE COLUMN STIMULUS AVERAGED OVER ALL OBSERVERS AND IMAGES

(MTF, LUMINANCE)

	(1,1)	(1,2)	(1,3)	(1,4)	(2,1)	(2,2)	(2,3)	(2,4)	(3,1)	(3,2)	(3,3)	(3,4)	(4,1)	(4,2)	(4,3)	(4,4)	(5,1)	(5,2)	(5,3)	(5,4)
(1,1)	.500	.292	.115	.240	.698	.594	.354	.448	.927	.875	.698	.552	.990	.979	.927	.833	.969	.969	.948	.958
(1,2)		.500	.458	.365	.823	.708	.646	.646	.990	.927	.948	.917	.979	.958	.979	.948	1.000	.990	.979	1.000
(1,3)			.500	.573	.927	.792	.802	.854	.896	.959	.979	.948	.990	.979	.990	1.000	1.000	.990	.990	.948
(1,4)				.500	.938	.781	.813	.635	.917	.927	.958	.927	.958	.906	.938	.969	.927	.979	1.000	.990
(2,1)					.500	.250	.240	.208	.719	.625	.469	.396	.917	.865	.656	.656	.938	.979	.958	.927
(2,2)						.500	.375	.458	.896	.802	.677	.740	.938	.969	.938	.906	.979	1.000	.958	.969
(2,3)							.500	.594	.948	.802	.750	.865	.917	.958	.948	.938	.969	.979	.990	.990
(2,4)								.500	.927	.813	.813	.844	.927	.875	.906	.938	.979	.958	.958	.990
(3,1)									.500	.354	.135	.188	.708	.635	.469	.385	1.000	.938	.875	.802
(3,2)										.500	.385	.417	.885	.708	.750	.771	.969	.979	.979	.969
(3,3)											.500	.552	.927	.813	.708	.813	.979	.990	.979	.948
(3,4)												.500	.885	.875	.760	.750	.948	.938	.958	.979
(4,1)													.500	.354	.260	.271	.813	.781	.656	.625
(4,2)														.500	.396	.427	.917	.833	.792	.771
(4,3)															.500	.667	.896	.885	.844	.875
(4,4)																.500	.865	.833	.906	.896
(5,1)																	.500	.365	.281	.271
(5,2)																		.500	.354	.406
(5,3)																			.500	.458
(5,4)																				.500

APPENDIX TABLE 2: PROBABILITY OF PREFERRING ROW STIMULUS OVER COLUMN STIMULUS FOR AIRPORT IMAGE AVERAGED OVER ALL OBSERVERS

(MTF, LUMINANCE)

	(1,1)	(1,2)	(1,3)	(1,4)	(2,1)	(2,2)	(2,3)	(2,4)	(3,1)	(3,2)	(3,3)	(3,4)	(4,1)	(4,2)	(4,3)	(4,4)	(5,1)	(5,2)	(5,3)	(5,4)
(1,1)	.500	.125	.104	.098	1.000	.472	.457	.409	1.000	.913	.826	.767	.979	1.000	1.000	.979	1.000	1.000	1.000	1.000
(1,2)		.500	.175	.372	1.000	1.000	.600	.674	1.000	.977	1.000	.938	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(1,3)			.500	.405	.933	.935	1.000	.822	1.000	1.000	1.000	1.000	1.000	.979	1.000	1.000	1.000	1.000	1.000	1.000
(1,4)				.500	.977	.872	.791	.957	.979	.936	.978	1.000	.979	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(2,1)					.500	.100	.200	.114	1.000	.634	.512	.488	.956	.978	.978	.884	1.000	1.000	1.000	1.000
(2,2)						.500	.275	.467	1.000	.974	.830	.911	1.000	1.000	.979	1.000	1.000	1.000	1.000	1.000
(2,3)							.500	.432	.978	.979	1.000	.911	.979	.979	1.000	1.000	.979	1.000	1.000	1.000
(2,4)								.500	.935	.889	.929	1.000	1.000	.979	.979	1.000	1.000	1.000	1.000	1.000
(3,1)									.500	.162	.143	.237	1.000	.684	.641	.707	1.000	1.000	1.000	1.000
(3,2)										.500	.341	.429	1.000	1.000	.881	.837	1.000	.979	1.000	1.000
(3,3)											.500	.568	1.000	.976	.900	.935	1.000	1.000	1.000	1.000
(3,4)												.500	.955	.860	.850	1.000	.979	1.000	1.000	.979
(4,1)													.500	.111	.118	.111	1.000	1.000	.968	.853
(4,2)														.500	.286	.414	1.000	1.000	.952	.949
(4,3)															.500	.600	1.000	1.000	1.000	.978
(4,4)																.500	1.000	1.000	1.000	1.000
(5,1)																	.500	.059	.000	.130
(5,2)																		.500	.133	.188
(5,3)																			.500	.188
(5,4)																				.500

APPENDIX TABLE 3: PROBABILITY OF PREFERRING ROW STIMULUS OVER COLUMN STIMULUS FOR CROP IMAGE AVERAGED OVER ALL OBSERVERS

(MTF, LUMINANCE)

	(1,1)	(1,2)	(1,3)	(1,4)	(2,1)	(2,2)	(2,3)	(2,4)	(3,1)	(3,2)	(3,3)	(3,4)	(4,1)	(4,2)	(4,3)	(4,4)	(5,1)	(5,2)	(5,3)	(5,4)
(1,1)	.500	.027	.021	.021	1.000	.395	.239	.149	1.000	.870	.553	.500	1.000	1.000	.875	.717	1.000	1.000	1.000	.978
(1,2)		.500	.195	.130	1.000	.933	.405	.457	1.000	.976	.979	.844	1.000	1.000	.957	1.000	1.000	1.000	1.000	1.000
(1,3)			.500	.333	1.000	.933	1.000	.750	1.000	1.000	1.000	.915	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(1,4)				.500	1.000	.936	.872	.957	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(2,1)					.500	.026	.042	.000	.952	.441	.250	.156	1.000	.878	.786	.650	1.000	1.000	1.000	.977
(2,2)						.500	.100	.178	1.000	.964	.644	.477	1.000	1.000	.957	.953	1.000	1.000	1.000	1.000
(2,3)							.500	.364	1.000	.977	.971	.818	1.000	1.000	1.000	.978	1.000	1.000	1.000	1.000
(2,4)								.500	1.000	1.000	1.000	.931	.979	1.000	1.000	1.000	.979	1.000	1.000	1.000
(3,1)									.500	.070	.000	.022	.689	.500	.222	.205	1.000	1.000	.943	.974
(3,2)										.500	.135	.182	1.000	.929	.727	.622	.979	1.000	1.000	.977
(3,3)											.500	.212	.979	.978	.968	.930	1.000	1.000	1.000	1.000
(3,4)												.500	.978	.935	.974	1.000	1.000	1.000	.979	1.000
(4,1)													.500	.059	.026	.077	.967	.808	.500	.696
(4,2)														.500	.091	.100	1.000	.972	.853	.939
(4,3)															.500	.281	.976	1.000	1.000	1.000
(4,4)																.500	1.000	.976	1.000	1.000
(5,1)																	.500	.000	.000	.048
(5,2)																		.500	.105	.100
(5,3)																			.500	.133
(5,4)																				.500

APPENDIX TABLE 4: PROBABILITY OF PREFERRING ROW STIMULUS
OVER COLUMN STIMULUS FOR MOUNTAIN IMAGE
AVERAGED OVER ALL OBSERVERS

(MTF, LUMINANCE)

	(1,1)	(1,2)	(1,3)	(1,4)	(2,1)	(2,2)	(2,3)	(2,4)	(3,1)	(3,2)	(3,3)	(3,4)	(4,1)	(4,2)	(4,3)	(4,4)	(5,1)	(5,2)	(5,3)	(5,4)
(1,1)	.500	.125	.063	.083	.792	.548	.366	.396	.955	.891	.826	.787	1.000	1.000	.957	.978	1.000	1.000	1.000	1.000
(1,2)		.500	.214	.292	.357	.935	.739	.771	1.000	1.000	.958	.936	1.000	.979	.979	.979	1.000	1.000	1.000	1.000
(1,3)			.500	.382	1.000	.955	1.000	.844	1.000	.979	1.000	.979	1.000	1.000	.979	1.000	1.000	1.000	1.000	1.000
(1,4)				.500	1.000	.957	.929	.969	1.000	.579	.979	1.000	.979	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(2,1)					.500	.026	.063	.087	.933	.605	.479	.500	1.000	.958	.933	.911	1.000	1.000	1.000	1.000
(2,2)						.500	.283	.205	.978	.975	.717	.867	1.000	.957	.979	.913	.979	.979	1.000	1.000
(2,3)							.500	.357	1.000	1.000	1.000	.909	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(2,4)								.500	.958	.979	.933	.971	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(3,1)									.500	.143	.071	.156	.897	.744	.829	.675	.978	.956	1.000	1.000
(3,2)										.500	.282	.375	.976	.973	.956	.884	1.000	1.000	1.000	1.000
(3,3)											.500	.447	.979	1.000	1.000	.927	1.000	1.000	.979	1.000
(3,4)												.500	.958	.952	1.000	1.000	1.000	1.000	1.000	1.000
(4,1)													.500	.034	.179	.220	1.000	.919	.912	.912
(4,2)														.500	.357	.452	1.000	1.000	.976	.976
(4,3)															.500	.484	1.000	1.000	1.000	1.000
(4,4)																.500	1.000	1.000	1.000	.975
(5,1)																	.500	.056	.105	.143
(5,2)																		.500	.278	.333
(5,3)																			.500	.300
(5,4)																				.500

APPENDIX TABLE 5: PROBABILITY OF PREFERRING ROW STIMULUS OVER COLUMN STIMULUS FOR OCEAN IMAGE AVERAGED OVER ALL OBSERVERS

(MTF,LUMINANCE)

	(1,1)	(1,2)	(1,3)	(1,4)	(2,1)	(2,2)	(2,3)	(2,4)	(3,1)	(3,2)	(3,3)	(3,4)	(4,1)	(4,2)	(4,3)	(4,4)	(5,1)	(5,2)	(5,3)	(5,4)
(1,1)	.500	.23[illegible]	.089	.222	.852	.615	.341	.444	.977	.909	.721	.558	1.000	1.000	.977	.881	1.000	.979	.978	.978
(1,2)		.500	.457	.356	.878	.857	.659	.652	1.000	1.000	.957	.935	1.000	1.000	1.000	.978	1.000	1.000	1.000	1.000
(1,3)			.500	.595	.936	.833	.914	.886	.913	.978	1.000	.978	1.000	.979	1.000	1.000	1.000	1.000	1.000	.978
(1,4)				.500	.977	.600	.875	.810	.955	.936	.978	.956	1.000	.933	.957	1.000	.956	1.000	1.000	1.000
(2,1)					.500	.184	.222	.182	.957	.643	.462	.361	1.000	.907	.674	.692	.977	1.000	.978	.956
(2,2)						.500	.357	.457	.932	.939	.707	.756	.977	1.000	.957	.953	1.000	1.000	1.000	1.000
(2,3)							.500	.615	.978	.854	.900	.927	.976	1.000	.978	.977	1.000	1.000	1.000	1.000
(2,4)								.500	.956	.857	.875	.946	.97[illegible]	.891	.933	1.000	1.000	1.000	.978	1.000
(3,1)									.500	.306	.093	.143	.955	.697	.462	.359	1.000	1.000	.950	.914
(3,2)										.500	.351	.405	.930	.857	.833	.795	1.000	1.000	1.000	1.000
(3,3)											.500	.576	.956	.857	.885	.857	1.000	1.000	1.000	.978
(3,4)												.500	.930	.929	.821	.900	1.000	1.000	[illegible]	1.000
(4,1)													.500	.250	.189	.194	1.000	.9[illegible]	[illegible]	.688
(4,2)														.500	.333	.394	.976	[illegible]	[illegible]	906
(4,3)															.500	.767	.975	[illegible]	[illegible]	[illegible]74
(4,4)																.500	.973	[illegible]	[illegible]	[illegible]00
(5,1)																	.500	[illegible]	[illegible]	42
(5,2)																		[illegible]	[illegible]	[illegible]286
(5,3)																			[illegible]	.375
(5,4)																				.500

APPENDIX TABLE 6: PROBABILITY OF PREFERRING ROW STIMULUS OVER COLUMN STIMULUS FOR PINES IMAGE AVERAGED OVER ALL OBSERVERS

(MTF, LUMINANCE)

	(1,1)	(1,2)	(1,3)	(1,4)	(2,1)	(2,2)	(2,3)	(2,4)	(3,1)	(3,2)	(3,3)	(3,4)	(4.1)	(4,2)	(4,3)	(4,4)	(5,1)	(5,2)	(5,3)	(5,4)
(1,1)	.500	.195	.295	.217	.962	.548	.500	.442	.939	.837	.732	.651	1.000	.911	.850	.821	1.000	.979	.978	1.000
(1,2)		.500	.568	.659	.864	.958	.844	.826	1.000	.975	.978	.932	.979	1.000	.957	1.000	1.000	1.000	1.000	1.000
(1,3)			.500	.750	.822	.659	1.000	.953	.956	.804	1.000	.957	1.000	.957	1.000	.979	1.000	.978	1.000	1.000
(1,4)				.500	.854	.477	.550	1.000	.905	.780	.778	1.000	1.000	.881	.872	1.000	1.000	.979	1.000	1.000
(2,1)					.500	.243	.275	.368	.875	.590	.683	.639	.973	.791	.725	.829	.976	1.000	.953	1.000
(2,2)						.500	.700	.705	.976	.920	.907	.902	.957	1.000	.978	.978	1.000	1.000	.958	1.000
(2,3)							.500	.610	.886	.780	.893	.889	.955	.955	1.000	1.000	.978	1.000	1.000	.979
(2,4)								.500	.905	.615	.537	1.000	.974	.854	.837	1.000	1.000	.977	1.000	1.000
(3,1)									.500	.300	.238	.289	1.000	.613	.625	.633	1.000	.946	.886	.941
(3,2)										.500	.472	.667	1.000	1.000	.842	.846	1.000	1.000	.955	.977
(3,3)											.500	.833	.974	.743	1.000	.951	1.000	1.000	1.000	1.000
(3,4)												.500	.829	.757	.559	.960	.927	1.000	1.000	1.000
(4,1)													.500	.107	.257	.300	.962	.839	.739	.632
(4,2)														.500	.516	.629	.947	1.000	.921	.972
(4,3)															.500	.677	1.000	.974	1.000	1.000
(4,4)																.500	.971	.886	.829	1.000
(5,1)																	.500	.059	.227	.368
(5,2)																		.500	.353	.611
(5,3)																			.500	.579
(5,4)																				.500

APPENDIX TABLE 7: PROBABILITY OF PREFERRING ROW STIMULUS
OVER COLUMN STIMULUS FOR OBSERVER EU
AVERAGED OVER ALL IMAGES

(MTF,LUMINANCE)

	(1,1)	(1,2)	(1,3)	(1,4)	(2,1)	(2,2)	(2,3)	(2,4)	(3,1)	(3,2)	(3,3)	(3,4)	(4,1)	(4,2)	(4,3)	(4,4)	(5,1)	(5,2)	(5,3)	(5,4)
(1,1)	.500	.128	.000	.025	.739	.306	.158	.075	.917	.700	.375	.359	1.000	.923	.737	.649	1.000	.975	.949	.974
(1,2)		.500	.231	.200	.950	.864	.410	.333	1.000	.972	.921	.769	1.000	.974	.897	.949	1.000	1.000	1.000	1.000
(1,3)			.500	.333	.975	.973	1.000	.794	1.000	1.000	1.000	.875	1.000	.975	1.000	1.000	1.000	1.000	1.000	1.000
(1,4)				.500	1.000	.974	.892	.870	1.000	1.000	1.000	1.000	.975	.975	.975	1.000	1.000	1.000	1.000	1.000
(2,1)					.500	.053	.050	.025	.903	.378	.150	.237	.947	.700	.564	.486	.950	1.000	.925	.947
(2,2)						.500	.231	.225	.975	.906	.550	.590	.975	.973	.850	.825	.975	1.000	.975	1.000
(2,3)							.500	.216	1.000	.950	.879	.730	1.000	.975	1.000	1.000	.975	1.000	1.000	.975
(2,4)								.500	.975	.925	.875	.889	1.000	1.000	.975	1.000	1.000	1.000	1.000	1.000
(3,1)									.500	.128	.050	.025	.880	.447	.359	.150	.974	.919	.838	.861
(3,2)										.500	.158	.225	.947	.941	.692	.538	1.000	1.000	.950	.974
(3,3)											.500	.405	.975	.900	.938	.923	1.000	1.000	.975	.975
(3,4)												.500	1.000	1.000	.895	.972	1.000	1.000	1.000	1.000
(4,1)													.500	.105	.075	.051	1.000	.763	.571	.429
(4,2)														.500	.229	.229	1.000	1.000	.838	.895
(4,3)															.500	.421	1.000	1.000	.972	1.000
(4,4)																.500	1.000	1.000	1.000	1.000
(5,1)																	.500	.069	.000	.032
(5,2)																		.500	.194	.226
(5,3)																			.500	.300
(5,4)																				.500

APPENDIX TABLE 8: PROBABILITY OF PREFERRING ROW STIMULUS
OVER COLUMN STIMULUS FOR OBSERVER BA
AVERAGED OVER ALL IMAGES

(MTF, LUMINANCE)

	(1,1)	(1,2)	(1,3)	(1,4)	(2,1)	(2,2)	(2,3)	(2,4)	(3,1)	(3,2)	(3,3)	(3,4)	(4,1)	(4,2)	(4,3)	(4,4)	(5,1)	(5,2)	(5,3)	(5,4)
(1,1)	500	.054	.050	.025	.970	.784	.486	.450	1.000	.975	.868	.769	1.000	1.000	.974	.919	1.000	1.000	1.000	.975
(1,2)		.500	.231	.051	1.000	.974	.784	.600	1.000	1.000	1.000	.974	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(1,3)			.500	.282	1.000	1.000	1.000	.895	1.000	1.000	1.000	.949	1.000	1.000	1.000	.975	1.000	1.000	1.000	1.000
(1,4)				.500	1.000	1.000	1.000	.975	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(2,1)					.500	.158	.051	.028	.973	.879	.579	.375	1.000	1.000	.829	.844	1.000	1.000	1.000	.975
(2,2)						.500	.158	.256	1.000	1.000	.895	.778	1.000	1.000	1.000	.974	1.000	.975	1.000	1.000
(2,3)							.500	.343	1.000	1.000	.975	.949	.975	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(2,4)								.500	.975	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(3,1)									.500	.194	.081	.026	1.000	.882	.452	.515	1.000	1.000	1.000	1.000
(3,2)										.500	.212	.125	1.000	.941	.941	.722	1.000	1.000	1.000	.974
(3,3)											.500	.258	.974	1.000	.974	.943	1.000	1.000	1.000	1.000
(3,4)												.500	1.000	.971	.950	.973	1.000	1.000	.975	1.000
(4,1)													.500	.107	.056	.057	1.000	.941	.818	.846
(4,2)														.500	.185	.148	.975	1.000	.946	.879
(4,3)															.500	.455	1.000	1.000	1.000	1.000
(4,4)																.500	1.000	.974	1.000	1.000
(5,1)																	.500	.111	.118	.087
(5,2)																		.500	.286	.182
(5,3)																			.500	.438
(5,4)																				.500

APPENDIX TABLE 9: PROBABILITY OF PREFERRING ROW STIMULUS OVER COLUMN STIMULUS FOR OBSERVER KK AVERAGED OVER ALL IMAGES

(MTF,LUMINANCE)	(1,1)	(1,2)	(1,3)	(1,4)	(2,1)	(2,2)	(2,3)	(2,4)	(3,1)	(3,2)	(3,3)	(3,4)	(4,1)	(4,2)	(4,3)	(4,4)	(5,1)	(5,2)	(5,3)	(5,4)
(1,1)	.500	100	.150	.200	1.000	.410	.350	.400	1.000	.925	.725	.600	1.000	.975	.950	.975	1.000	1.000	1.000	1.000
(1,2)		.500	.350	.400	.897	.875	.500	.850	1.000	1.000	.975	.950	.975	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(1,3)			.500	.545	.875	.769	.875	.886	.950	.846	1.000	.974	1.000	.950	1.000	1.000	1.000	1.000	1.000	1.000
(1,4)				.500	.900	.725	.641	1.000	.900	.825	.800	1.000	1.000	.875	.925	1.000	1.000	1.000	1.000	1.000
(2,1)					.500	.132	.175	.200	.909	.389	.525	.475	1.000	.949	.950	.769	1.000	1.000	1.000	1.000
(2,2)						.500	.300	.375	.950	1.000	.800	.875	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(2,3)							.500	.622	.950	.825	1.000	.886	.975	.974	1.000	.974	1.000	1.000	1.000	1.000
(2,4)								.500	.900	.775	.618	1.000	.975	.875	.872	1.000	1.000	.975	1.000	1.000
(3,1)									.500	.179	.128	.275	.938	.625	.625	.625	1.000	1.000	.950	1.000
(3,2)										.500	.425	.575	1.000	1.000	.825	.875	1.000	1.000	1.000	1.000
(3,3)											.500	.647	1.000	.846	.818	.919	1.000	1.000	1.000	1.000
(3,4)												.500	.846	.700	.703	1.000	.950	1.000	1.000	1.000
(4,1)													.500	.103	.250	.385	1.000	.900	.825	.846
(4,2)														.500	.359	.525	1.000	1.000	.900	1.000
(4,3)															.500	.600	.975	.974	1.000	.974
(4,4)																.500	.975	.900	.895	1.000
(5,1)																	.500	.029	.100	.132
(5,2)																		.500	.150	.275
(5,3)																			.500	.257
(5,4)																				.500

APPENDIX TABLE 10: PROBABILITY OF PREFERRING ROW STIMULUS OVER COLUMN STIMULUS FOR OBSERVER JB AVERAGED OVER ALL IMAGES

(MTF, LUMINANCE)

	(1,1)	(1,2)	(1,3)	(1,4)	(2,1)	(2,2)	(2,3)	(2,4)	(3,1)	(3,2)	(3,3)	(3,4)	(4,1)	(4,2)	(4,3)	(4,4)	(5,1)	(5,2)	(5,3)	(5,4)
(1,1)	.500	.054	.100	.050	.929	.270	.179	.282	1.000	.846	.692	.658	1.000	1.000	1.000	.895	1.000	1.000	1.000	1.000
(1,2)		.500	.125	.200	.974	.923	.450	.550	1.000	1.000	1.000	.875	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(1,3)			.500	.444	.975	.900	.944	.692	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(1,4)				.500	.974	.850	.865	1.000	.975	.925	.975	1.000	1.000	1.000	.975	1.000	.975	1.000	1.000	1.000
(2,1)					.500	.027	.100	.128	1.000	.375	.389	.405	1.000	.857	.714	.857	1.000	1.000	1.000	1.000
(2,2)						.500	.308	.250	1.000	1.000	.605	.641	1.000	1.000	.974	1.000	1.000	1.000	.974	1.000
(2,3)							.500	.361	.974	.949	1.000	.897	1.000	1.000	1.000	.975	1.000	1.000	1.000	1.000
(2,4)								.500	.975	.923	.921	1.000	1.000	.974	.974	1.000	1.000	1.000	1.000	1.000
(3,1)									.500	.086	.075	.135	.929	.400	.552	.520	1.000	.970	1.000	1.000
(3,2)										.500	.189	.300	1.000	1.000	.818	.730	1.000	1.000	1.000	1.000
(3,3)											.500	.417	.947	.974	1.000	.833	1.000	1.000	1.000	1.000
(3,4)												.500	.914	.946	.811	1.000	1.000	1.000	1.000	.975
(4,1)													.500	.000	.067	.125	1.000	.889	1.000	.900
(4,2)														.500	.333	.346	1.000	.970	.950	1.000
(4,3)															.500	.552	1.000	1.000	1.000	1.000
(4,4)																.500	1.000	1.000	1.000	1.000
(5,1)																	.500	.000	.143	.000
(5,2)																		.500	.000	.500
(5,3)																			.500	1.000
(5,4)																				.500

APPENDIX TABLE 11: PROBABILITY OF PREFERRING ROW STIMULUS OVER COLUMN STIMULUS FOR OBSERVER MB AVERAGED OVER ALL IMAGES

(MTF, LUMINANCE)

	(1,1)	(1,2)	(1,3)	(1,4)	(2,1)	(2,2)	(2,3)	(2,4)	(3,1)	(3,2)	(3,3)	(3,4)	(4,1)	(4,2)	(4,3)	(4,4)	(5,1)	(5,2)	(5,3)	(5,4)
(1,1)	.500	.267	.242	.241	.895	.609	.500	.464	.929	.900	.885	.741	.973	1.000	.933	.815	1.000	.974	1.000	1.000
(1,2)		.500	.515	.781	.818	.882	.912	.813	1.000	.941	.946	.941	1.000	1.000	.973	1.000	1.000	1.000	1.000	1.000
(1,3)			.500	.793	.882	.625	.966	.861	.914	.865	1.000	1.000	1.000	.974	.974	1.000	1.000	.974	1.000	.973
(1,4)				.500	.923	.543	.571	.842	.931	.813	.875	.946	.966	.935	.919	1.000	1.000	.974	1.000	1.000
(2,1)					.500	.148	.286	.167	.857	.654	.522	.368	.964	.935	.926	.909	1.000	1.000	1.000	1.000
(2,2)						.500	.581	.655	.931	.893	.788	.821	.939	.973	1.000	.968	1.000	1.000	1.000	1.000
(2,3)							.500	.727	.912	.829	.929	.879	.926	.970	.973	1.000	.971	1.000	1.000	1.000
(2,4)								.500	.929	.655	.750	.969	.926	.793	.912	1.000	.969	1.000	.971	1.000
(3,1)									.500	.400	.200	.333	1.000	.773	.538	.647	1.000	1.000	1.000	.941
(3,2)										.500	.538	.708	.920	.875	.857	.963	.970	1.000	1.000	1.000
(3,3)											.500	.846	.963	.846	.917	.938	1.000	1.000	1.000	1.000
(3,4)												.500	.815	.720	.792	.917	.966	1.000	.943	1.000
(4,1)													.500	.357	.400	.368	.833	1.000	1.000	1.000
(4,2)														.500	.476	.500	.905	.950	.957	.938
(4,3)															.500	.696	.933	.950	1.000	.957
(4,4)																.500	1.000	.882	.917	.952
(5,1)																	.500	.500	.200	.800
(5,2)																		.500	.500	.800
(5,3)																			.500	.500
(5,4)																				.500

APPENDIX TABLE 12: PROBABILITY OF PREFERRING ROW STIMULUS OVER COLUMN STIMULUS FOR OBSERVER MD AVERAGED OVER ALL IMAGES

(MTF,LUMINANCE)

	(1,1)	(1,2)	(1,3)	(1,4)	(2,1)	(2,2)	(2,3)	(2,4)	(3,1)	(3,2)	(3,3)	(3,4)	(4,1)	(4,2)	(4,3)	(4,4)	(5,1)	(5,2)	(5,3)	(5,4)
(1,1)	.500	.538	.154	.256	1.000	.880	.658	.550	1.000	.972	.900	.821	1.000	1.000	1.000	.975	1.000	1.000	1.000	1.000
(1,2)		.500	.727	.629	1.000	1.000	.969	.947	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(1,3)			.500	.889	.923	.879	1.000	1.000	.975	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(1,4)				.500	.975	.750	.875	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.975	1.000	1.000	1.000
(2,1)					.500	.267	.308	.308	1.000	.933	.711	.632	1.000	1.000	.975	.973	1.000	1.000	1.000	1.000
(2,2)						.500	.667	.750	1.000	.967	.970	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(2,3)							.500	.857	.974	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(2,4)								.500	.950	.935	.960	1.000	1.000	1.000	.975	1.000	1.000	1.000	1.000	1.000
(3,1)									.500	.250	.167	.273	.933	.882	.828	.735	1.000	1.000	1.000	1.000
(3,2)										.500	.563	.613	1.000	1.000	1.000	1.000	1.000	.975	1.000	1.000
(3,3)											.500	.923	1.000	.917	1.000	.971	1.000	1.000	1.000	1.000
(3,4)												.500	1.000	.970	1.000	1.000	.974	1.000	1.000	1.000
(4,1)													.500	.000	.217	.115	1.000	1.000	.963	.958
(4,2)														.500	.429	.650	1.000	1.000	1.000	1.000
(4,3)															.500	1.000	1.000	1.000	1.000	1.000
(4,4)																.500	.971	1.000	1.000	1.000
(5,1)																	.500	.000	.000	.333
(5,2)																		.500	.667	1.000
(5,3)																			.500	.500
(5,4)																				.500

☆U.S. GOVERNMENT PRINTING OFFICE: 1994 - 560-051/00019